Methods to Improve the Prediction Accuracy and Performance of Ensemble Models

Vincent 'Fisayo Adegoke

https://orcid.org/0000-0001-6848-9840

Division of Computer Science and Informatics

School of Engineering

A thesis submitted in partial fulfilment of the requirements of London South Bank University for the degree of Master of Philosophy

September 2021

Abstract

The application of ensemble predictive models has been an important research area in predicting medical diagnostics, engineering diagnostics, and other related smart devices and related technologies. Most of the current predictive models are complex and not reliable despite numerous efforts in the past by the research community. The performance accuracy of the predictive models have not always been realised due to many factors such as complexity and class imbalance. Therefore there is a need to improve the predictive accuracy of current ensemble models and to enhance their applications and reliability and non-visual predictive tools.

The research work presented in this thesis has adopted a pragmatic phased approach to propose and develop new ensemble models using multiple methods and validated the methods through rigorous testing and implementation in different phases. The first phase comprises of empirical investigations on standalone and ensemble algorithms that were carried out to ascertain their performance effects on complexity and simplicity of the classifiers. The second phase comprises of an improved ensemble model based on the integration of Extended Kalman Filter (EKF), Radial Basis Function Network (RBFN) and AdaBoost algorithms. The third phase comprises of an extended model based on early stop concepts, AdaBoost algorithm, and statistical performance of the training samples to minimize overfitting performance of the proposed model. The fourth phase comprises of an enhanced analytical multivariate logistic regression predictive model developed to minimize the complexity and improve prediction accuracy of logistic regression model.

To facilitate the practical application of the proposed models; an ensemble non-invasive analytical tool is proposed and developed. The tool links the gap between theoretical concepts and practical application of theories to predict breast cancer survivability.

The empirical findings suggested that: (1) increasing the complexity and topology of algorithms does not necessarily lead to a better algorithmic performance, (2) boosting by resampling performs slightly better than boosting by reweighting, (3) the prediction accuracy of the proposed ensemble EKF-RBFN-AdaBoost model performed better than several established ensemble models, (4) the proposed early stopped model converges faster and minimizes overfitting better compare with other models, (5) the proposed multivariate logistic regression concept minimizes the complexity models (6) the performance of the proposed analytical non-invasive tool performed comparatively better than many of the benchmark analytical tools used in predicting breast cancers and diabetics ailments.

The research contributions to ensemble practice are: (1) the integration and development of EKF, RBFN and AdaBoost algorithms as an ensemble model, (2) the development and validation of ensemble model based on early stop concepts, AdaBoost, and statistical concepts of the training samples, (3) the development and validation of predictive logistic regression model based on breast cancer, and (4) the development and validation of a non-invasive breast cancer analytic tools based on the proposed and developed predictive models in this thesis.

To validate prediction accuracy of ensemble models, in this thesis the proposed models were applied in modelling breast cancer survivability and diabetics' diagnostic tasks. In comparison with other established models the simulation results of the models showed improved predictive accuracy.

The research outlines the benefits of the proposed models, whilst proposes new directions for future work that could further extend and improve the proposed models discussed in this thesis.

Dedication

I dedicate this thesis to the memory of my beloved parents in gratitude. I also dedicate this work to my adored wife Catherine who encourage me to pursue my dreams and finish my dissertation and to my children Praise, Precious and Princess for their understanding. This dissertation is also dedicated to my senior brother Professor G.0 Adegoke for his active role in my life.

Acknowledgements

I would like to express my sincere appreciation to my supervisors: Dr Daqing Chen, Dr Safia Barikzai, Professor Dilip Patel and Professor Ebad Banissi for their invaluable assistance and support throughout the research of this thesis.

In addition, I am also very grateful to John Harper and Louise Thompson for their administrative support and advice during my research at the London South Bank University (LSBU). I am also grateful to Michael Damiani my Line Manager for his understanding and support during the programme. Thanks also to Patrice Seuwou and other colleagues at the LSBU for their support.

The achievements of this thesis would not have been possible without the support and understanding of my wife, Catherine and my children: Praise, Precious and Princess during this study. I would eternally be grateful to my senior brother Professor G.O Adegoke who took the mantle of parents, worked hard and gave me happy life.

Above all I would forever be grateful to the Almighty God for His Love and Grace over my life and my family.

Declaration

I hereby declare that I have produced this thesis and that it has not been submitted anywhere else for any previous award. I did not receive assistance or aids from any third parties, it was a self-sponsored research.

The collaborative notions taken directly or indirectly from other sources have been indicated and clearly acknowledged. Some of the work derived from this thesis have been accepted and published:

1. Adegoke, V., Banissi, E. & Barikzai, S., 2019. Improving Prediction Accuracy of Breast Cancer Survivability and Diabetes Diagnosis via RBF Networks trained with EKF models. International Journal of Computer Information Systems and Industrial Management Application, 11(2019), pp. 082-100.

2. Seuwou, Patrice and Vincent F. Adegoke. "The Changing Global Landscape with Emerging Technologies and Their Implications for Smart Societies." Handbook of Research on 5G Networks and Advancements in Computing, Electronics, and Electrical Engineering, edited by Augustine O. Nwajana and Isibor Kennedy Ihianle, IGI Global, 2021, pp. 402-423.

3. Adegoke, V., Chen, D., Banissi, E. & Barikzai, S., 2019. Enhancing Ensemble Prediction Accuracy of Breast Cancer Survivability and Diabetes Diagnostic Using Optimized EKF-RBFN Trained Prototypes. In: Madureira A., Abraham A., Gandhi N., Silva C., Antunes M. (eds.) Proceedings of the Tenth International Conference on Soft Computing and Pattern Recognition (SoCPaR 2018) Advances in Intelligent Systems and Computing, vol. 942. Springer, Cham. https://doi.org/10.1007/978-3-030-17065-3_6

4. Adegoke, V., Chen, D., Banissi, E., and Barikzai, S. "Prediction of breast cancer survivability using ensemble algorithms," IEEE *2017 International Conference on Smart Systems and Technologies (SST)*, 2017, pp. 223-231, doi: 10.1109/SST.2017.8188699.

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Acronyms

AUC	Area Under Curve
BI	Business Intelligent
BBM	Boost by Majority
EKF	Extended Kalman filter
EKF-RBFN	RBFN models trained with EKF
EKF-RBFN-AdaBoost	Ensembles of RBFN models trained with EKF
FNR	False Negative Rate
FPR	False Positive Rate
GA	Genetic Algorithm
GA-SVM	Hybrid of GA and SVM
GBM	Gradient Boosting Machines
GD	Gradient Descent
GD-RBFN	RBFN models trained with GD
GD-RBFN-AdaBoost	Ensembles of RBFN models trained with GD
GUI	Graphical User Interface
IDE	Integrated Development Environment
KF	Kalman Filter
KL	Kullback-Leibler Divergence
MAE	Mean Absolute Error
LAS	London Ambulance Service, NHS Trust
LQE	Linear Quadratic Estimate
MATLAB	Matrix Laboratory
MART	Multiple Additive Regression Tree
MLPNN	Multi-Layer Perceptron Neural Networks
MLP	Multi-Layer Perceptron

NN	Neural Networks
PSO	Particle Swarm Optimization
PRC	Precision Recall curves
RBFN	Radial Basis Function Network
RMSLE	Root Mean Squared Log Error
ROC	Receiver Operator Characteristic curve
SVM	Support Vector Machine
TNR	True Negative Rate
UCI	UC Irvine Machine Learning Repository
UKF	Unscented Kalman Filter
UPF	Unscented Particle Filter
TPR	True Positive Rate
WHO	World Health Organization
VAR	Vector Auto Regression

Chapter 1: Introduction

1.1 Research Outline

This chapter introduces the proposed research and outlines the research background by providing an introduction to the research study undertaken. The study examines the current problems of ensemble models and methods to enhance its prediction and performance accuracy, a topic that is relatively unexplored in the literature.

The rest of this chapter is organised as follows. 1.2 provides discussion on background to ensemble prediction, importance of ensemble concepts and the research problem. Section 1.3 states the research aims, objectives and the research hypothesis. The research approach and methodology are presented in Section 1.4. Original contributions to knowledge and publications resulting from this thesis are presented in Section 1.5. The structure of the thesis is presented and discussed in Section 1.6. Finally, Section 1.7 concludes the chapter with a summary.

1.2 Research Problem

1.2.1 Background of Ensemble Predictive Methods

The knowledge of predictive technology is not new, Cogburn states that the ancient Chinese farmers created the first predictive solar calendar to forecast climate changes (Cogburn, 2019). Likewise, in 1940s the British intelligence used advanced predictive techniques (Gladwin, 1997) to break the German Machine Ciphers in World War II. The emergence of big data, AI and machine learning technologies has triggered the need to identify trends, and to accurately predict potential problems. At the same time it has also increased the reliance on AIs and predictive technology to carry out tasks at a greater efficiency and accuracy. This has enabled researchers to understand, to monitor and to optimize predictive processes and

efficient linguistic decision making for robotic route learning (He, et al., 2014). The complexity of devices that depend on ensemble predictive methods and simulation models in monitoring life-dependent and safety-critical systems in the medical and industrial sectors is increasing and becoming more demanding (Ishak & Tokhi, 2017). Its predictive accuracy and reliability can no longer be ignored. Therefore, there is a need for improved ensemble methods in order to maximize the potentials of business intelligence and analytics information that can be derived from the Big Data using contemporary data mining methods. Doing this will facilitate the development of diagnostic tools that depends on effective performance of the predictive ensemble methods.

1.2.2 Importance of Ensemble Predictive Methods

Failure of predictive methods and their poor performances that result in fatal accidents and misdiagnosis are common problems in industries and health sectors (Zhang, et al., 2019; Alharthi, 2018). The way algorithms are designed and trained plays major roles in machine learning predictive accuracy (Walker, et al., 2020; Wang, 2008) and their reliability. Some of the numerous factors affecting the performance of ensemble methods among others are the inter-classifiers relationship, weight training and updates, combining methods, diversity generator and the ensemble size (Kuncheva, 2014) of the members. However, many of these have not been fully considered in ensemble development in particular AdaBoost an ensemble technique.

The AdaBoost (that is Adaptive Boosting) algorithm (Freund & Schapire, 1977) is a statistical classification meta-algorithm that be used in combination with other learning algorithms to boost their predictive performance.

AdaBoost has many potential applications and has been applied in many areas such as text classification, natural language processing, drug discovery and computational biology (Fan, et al., 2015) vision and object recognition (Viola & Jones, 2004; Lee, et al., 2013), medical diagnosis (Abuhasel, et al., 2015) and industrial chemical fault diagnosis (Karimi & Jazayeri-Rad, 2014), however with predictive accuracy issues and other performance challenges.

AdaBoost like other ensemble algorithms is primarily based on iterative training of the base classifiers on the samples.

However, recent studies show that their predictive outputs are currently inconsistent (Jeon, et al., 2019; Sun, et al., 2006) due to weaknesses such that its predictive accuracy is not as it should be.

Therefore, there is a need to develop improved models that meet the expected predictive output and are acceptable to the end users and other stakeholders.

1.2.3 Research Problem and Gaps

AdaBoost has two implementation methods, these are boosting by sampling and boosting by reweighting. Souza and Matwin (2012), argues that it is unclear which of these two methods performs better (Souza & Matwin, 2012; Seiffert, et al., 2008) therefore there is a need to find out which of the methods performs better and enhance the predictive accuracy of the algorithms (Wang, 2008; Freund & Schapire, 2014). This will allow to determine which implementation method could improve overall predictive performance of the AdaBoost technique. This could also provide performance breakdown of the implementation methods and further direction on which of the two methods could be enhanced to increase predictive performance of AdaBoost technique. Even though several standalone algorithms have proved to be impressive in many applications. However, study shows that there are some situations

where standalone methods might not be able to produce the required predictive results (Lee, 2018) when handling complicated tasks and applications that predictive accuracy is essential. The advantage of predictive ensemble methods is essential for predictive tasks that requires high likelihoods accuracy such as in non-invasive breast cancer analytical tool as addressed in this study.

EKF (Lima, et al., 2017; Chernodub, 2014), RBFN (Wang, et al., 2006; Oliveira, 2012) and AdaBoost (Freund & Schapire, 2014) models have been used extensively in predictive areas and classification tasks with promising results. However, reviews show that no research has been conducted to bridge the three algorithms together as an ensemble model to address current: predictive accuracy, class imbalance and overfitting problems of ensemble methods.

Therefore, one of the goals of this study is to investigate the prediction performance of training RBFN (Wang, 2017; Zhang & Yang, 2013) with EKF (Kamath, et al., 2011; Krok, 2013), then apply the trained networks model as a base classifier with ensemble AdaBoost (Lin & Wang, 2011; Karimi & Jazayeri-Rad, 2014) to improve the predictive performance of the model. Doing this would enable AdaBoost to generate and combine several weak classifiers that produces a stronger predictive output. Additionally, the Extended Kalman Filter (EKF) would serve as an optimizing agent to train the Radial Basis Function Network (RBFN) parameters.

1.3 Research Aims and Objectives

Reviews show that, contemporary developments in predictive technology coupled with advanced language processing and deeper artificial intelligence can help to forecast and mitigate infectious diseases such as Covid-19 outbreak, flu epidemic (GT, 2020; Imhoff & Lamberty, 2020). Therefore, application of predictive technology could promote smart living

society (Seuwou & Adegoke, 2021) and supplement environmental sustainability. However, relatively few studies have been focused on enhancing ensemble predictive accuracy and non-invasive analytical diagnostic tools that can enhance the adoption of the techniques. Therefore, in this research the focus is to investigate, propose extended ensemble models to enhance the performance accuracy of ensemble methods. The specific objectives of this study are:

1. **Objective 1** - To review and carry out empirical investigations in comparing performance of boosting methods, predictive performance of ensemble and standalone predictive techniques. The investigation outcome will assist in understanding the current performance issues of predictive methods and possible ways of addressing them. It will also provide answer to the following hypothesis:

   Hypothesis 1 – Boosting implementation method of AdaBoost: boosting by reweighting or boosting by resampling is positively connected to the predictive accuracy of the technique.

   Hypothesis 2 – The predictive performance of ensemble AdaBoost method is positively linked to the complexity and topology of the chosen classifier.

2. **Objective 2** - To propose and develop a new ensemble model by integrating Extended Kalman Filter (EKF) in training Radial Basis Function Networks (RBFN) to optimize the parameters of the network. Then apply the trained network prototypes (EKF- RBFN) as a base classifier with AdaBoost as a meta-model to enhance stronger predictions.
3. **Objective 3** To propose and develop an ensemble model based on early stopping concept, and statistics data of the training samples namely the mean, standard deviation and

thresholds in order to minimise generalization error and to avoid overtraining that may result in overfitting.

4. **Objective 4** To propose and develop ensemble model based on the concepts of ensemble multivariate logistic regression that best predicts the binary response variable Y for the values of multiple X variables of the predictors.
5. **Objective 5** To develop analytical tools that combine the theoretical concepts and practical application of the proposed models as non-invasive breast cancer survivability and dietetics diagnostic predictive devices.
6. **Objective 6** To validate the performance of the proposed models using benchmark samples and to suggest recommendations on best approach for future work.

1.4 Research Methodology

For the purpose of accomplishing the above stated objectives, the research was divided into four phases and conducted accordingly as summarised below:

1. **Phase One -**The main purpose of this phase is to achieve Objective 1 and are in two stages. Firstly, to resolve boosting methods performance discrepancies in the literature. Secondly, to compare performance efficiencies of ensemble and standalone models in order to provide answers to Hypothesis 1 and Hypothesis 2. Doing this will enable better understanding of the boosting methods, standalone and ensemble techniques and how to improve their performance.
2. **Phase Two -** The emphasis of this phase is to improve the prediction accuracy of AdaBoost as an ensemble model by integrating EKF, RBFN and ensemble AdaBoost as meta-classifier to achieve the purpose of Objective 2 of this study.

3. **Phase Three -** The focus of this phase is to realise Objective 3, to improve the performance of AdaBoost by applying early stopping notions and the statistical performance of the training sample to overcome overfitting and generalization issues.
4. **Phase Four –** To achieve Objective 4, this phase is implemented in two parts, the first part is to identify breast cancer prognostic feature and develop a logistic regression model using multivariate features of the samples. The second part of this phase is to develop a visual non-invasive breast cancer analytical tool in order to accomplish Objective 5. The tool makes exclusive use of the proposed model to show how the proposed models can be translated into practice to facilitate the application and evaluation of the proposed models and focuses on the empirical testing and analysis of the proposed models and assessment of the analytical tool.

To accomplish Objective 6, at each phase of the study the performance of the proposed models were tested and validated using benchmark samples. To accomplish Objective 7 recommendations and suggestions for future research were also provided

1.5 Original Contributions to Knowledge

The key contributions of this work are:

I. **Extended 3-in-1 ensemble predictive model.** This model integrates EKF- RBFN prototypes and ensemble AdaBoost as a meta-classifier to improve the performance of ensemble models.
II. **Extended model based on early stopping concepts.** This model using the statistical performance and thresholds of the training models to minimize overfitting problem in ensemble models.

III. **Development of Logistic regression approach model.** The model was based on multivariate concept to develop a predictive breast cancer logistic model.

IV. **Development of analytical breast cancer and diabetics diagnostic tools.** Using the concepts of developed ensemble models developed in this thesis. The developed tool is cost-effective, non-invasive that could be used to predict the breast cancer and diabetics diagnostics.

V. **Hypothesis and Boosting Methods**. The study has confirmed that the complication and the topology (the computational geometry and computational complexity) of algorithms does not necessarily improve the performance accuracy of algorithmic methods. The study further indicates that implementing AdaBoost by resampling method performs slightly better than implementing it by reweighting method.

In addition, the proposed models have been tested and validated. The prediction performances were better than some of the benchmark models. The study has achieved the objective of developing ensemble models that improved the prediction accuracy of existing models.

Publications Resulting from the Thesis

The research outcomes to date have been published in peer reviewed conferences, journals and book chapters, see below:

Journal Publications

1. Adegoke, V., Banissi, E. & Barikzai, S., 2019. Improving Prediction Accuracy of Breast Cancer Survivability and Diabetes Diagnosis via RBF Networks trained with EKF models.

International Journal of Computer Information Systems and Industrial Management Application, 11(2019), pp. 082-100.

Book Chapters

1. Seuwou, Patrice and Vincent F. Adegoke. "The Changing Global Landscape with Emerging Technologies and Their Implications for Smart Societies." Handbook of Research on 5G Networks and Advancements in Computing, Electronics, and Electrical Engineering, edited by Augustine O. Nwajana and Isibor Kennedy Ihianle, IGI Global, 2021, pp. 402-423.
2. Adegoke, V., Chen, D., Banissi, E. & Barikzai, S., 2019. -Enhancing Ensemble Prediction Accuracy of Breast Cancer Survivability and Diabetes Diagnostic Using Optimized EKF-RBFN Trained Prototypes. In: Madureira A., Abraham A., Gandhi N., Silva C., Antunes M. (eds.) Proceedings of the Tenth International Conference on Soft Computing and Pattern Recognition (SoCPaR 2018) Advances in Intelligent Systems and Computing, vol. 942. Springer, Cham. https://doi.org/10.1007/978-3-030-17065-3_6

International Conference Publications

1. Adegoke, V., Chen, D., Banissi, E., and Barikzai, S. "Prediction of breast cancer survivability using ensemble algorithms," IEEE *2017 International Conference on Smart Systems and Technologies (SST)*, 2017, pp. 223-231, doi: 10.1109/SST.2017.8188699.
2. Adegoke, V., Chen, D., Banissi, E. & Barikzai, S., 2017. IEEE *Predictive Ensemble Modelling: An Experimental Comparison of Boosting Implementation Methods. European Modelling Symposium on Mathematical modelling and Computer simulation,* Manchester, UK. 20-22 November 2017.

The Google scholar url for above publications is

https://scholar.google.com/citations?hl=en&user=Igh_IGQAAAAJ

1.6 Thesis Structure

In order to achieve the aims and objectives of the research, this thesis comprises of seven chapters which is outlined here.

Chapter 1 provides a brief introduction to the background, rationale and objectives of the study. The chapter also outlines the contributions, publications resulting from the research and structure of this thesis.

Chapter 2, presents a generic and critical literature of some of the relevant research work that are related to ensemble modelling. It also states the research hypothesis, in provides and identify reasonable research gaps in the literature.

Chapter 3 focuses on empirical investigations: ensemble boosting methods, standalone and ensemble models. It also validates and provide answers to Hypothesis 1 and Hypothesis 2 stated in chapter 2 of this thesis.

Chapter 4, proposes a new algorithm framework named LSB-EKF-RBFN-AdaBoost in order to enhance the predictive performance of ensemble model. In the framework, EKF is applied to train and optimize the training parameters of Radial Basis Function Networks (RBFN) to generate optimized EKF-RBFN prototypes. The optimized prototypes are then applied as base classifiers with AdaBoost as an ensemble meta-algorithm to train and combined prediction as final output. The simulation results are presented and discussed.

Chapter 5 explores how to enhance ensemble prediction accuracy, and proposes two algorithms that were based on early stopping and logistic regression concepts respectively.

The research also proposes a non-invasive analytic model for the predicting of breast cancer in patients. The framework was integrated with the predictive models developed in this thesis using benchmarked Wisconsin breast cancer survivability datasets.

Chapter 6 presents the discussion and conclusion of the thesis. It also provides some of the main benefits, limitations of the proposed models, methods presented in this thesis, and recommendations when to apply the models.

Chapter 7 presents the summary of the research and concluding remarks of the results of the models presented in this thesis, and finally propose prospective directions for future research.

1.7 Chapter Summary

The discussion in this chapter has elucidated the background of this thesis. The aim of this study was to improve the performance output of predictive ensemble techniques by developing new ensemble models and non-invasive visual diagnostic tool based on the developed models. This chapter has provided an introduction to the research problems and gaps. It further establish the focus of this study. The research methodology and research contributions of the study in various phases have been briefly discussed and analysed. The resulting publications from the research have been provided. Finally, the structure and brief description of all the chapters of this thesis have also discussed and presented. The next chapter will provide literature review on ensemble techniques and related technologies in order to identify research gaps.

Chapter 2: Literature Review and Gaps in Knowledge

2.1 Introduction

The main purpose of ensemble algorithms is to integrate and combine the predictions of multiple trained models in order to obtain a highly accurate classification and prediction rules. The final prediction is based on the combinational output of a set of participating learners instead of relying on a single classifier. There are two paradigms of ensemble methods, these are sequential ensemble methods where the base learners are generated sequentially and parallel ensemble methods where the base learners are generated in parallel.

This is because ensemble model combines multiple (mainly diverse) models together to deliver more stable and predict better than a single model. The concept is synonymous to seeking the opinion of experts before taking a final decision in responding or resolving a problem at hand (Polikar, 2006).

Ensemble methods create multiple models then combine the decision boundaries of the participating members or committees to produce more accurate and improved predictive results.

Despite the fact that there are numerous interesting research works in ensemble methods, reviews show that many of the variants of ensemble methods are still limited in performance efficiency, prediction accuracy and computational overhead requirements than expected due to the complexity (time and space) and topology of the algorithms.

The complexity of an algorithm is a measure of amount of time, and memory or space describing the efficiency of the algorithm in terms of the amount of data or size the algorithm must process (Jimenez, 1998; Gao & Xu, 2014). Complexity efficiency can be measured in

terms of time and space. Time complexity is a function of the amount of time an algorithm takes in terms of the amount of input to the algorithm. Space complexity is a function of the amount of memory an algorithm takes in terms of the amount of input to the algorithm. Hence, one of the main objectives of this research is to develop simple and non-complex ensemble models with few nodes that can perform more effectively than current methods.

Due to the complexities and task requirements in ensemble predictions, researchers have come up with a variety of models and theories to explain their performance and adoption. Therefore, it is important to purposefully study them, their theoretical concepts and implications. This would provide sound knowledge of current methods, their limitations, and problems and how it could be improved for better predictive performance.

Therefore, reducing the current issues such as overfitting, complexity and improving prediction performance of current models.

The structure of this chapter is as follows. **Section 2.2** provides a review of current state of ensemble models. It also explores the limitations and potential benefits of ensemble models. **Section 2.3** is a brief chronological timeline of ensemble models. **Section 2.4** provides a review of the concepts and theories underlying RFFN and EKF techniques. **Section 2.5** discusses identified gaps in the literature. **Section 2.6** provides the evaluation criteria of ensemble models. **Section 2.7** provides ensemble class imbalance. **Section 2.8** discusses ensemble diversity. **Section 2.9** presents ensemble classifier outputs and **Section 2.10** provides summary of the chapter.

2.2 Ensemble Models

The primary purpose of ensemble methods in classification tasks are to improve the prediction accuracy of models by producing a stronger output model through the

combination of multiple weak models. The classification process of ensemble techniques is in two stages: the selection and training of the base classifiers, and the process of combining the weak classifiers to form a stronger classifier better than guessing. It has been proven that the concept boosts the overall performance of ensemble models by taking the dependency advantage and diversity among the various base learners to form a stronger classifier (Freund & Schapire, 2014).

The idea is as graphically illustrated in Figure 2-1, it shows overlapping training sample subsets that are randomly selected to train 3-classifiers: Model 1, Model 2 and Model 3. The classifiers are trained and the three diverse decision boundaries are then combined to obtain a more accurate classification.

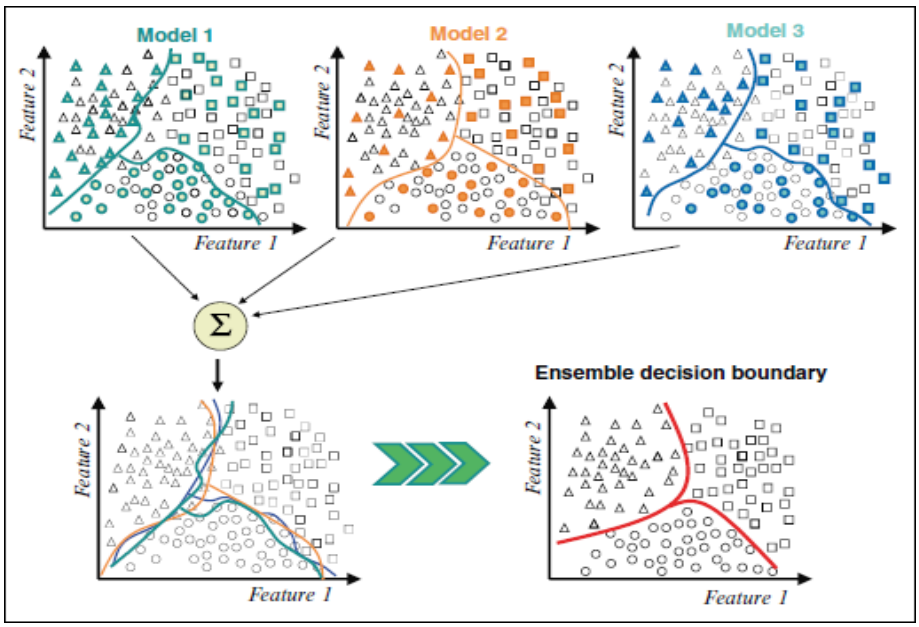


Figure 2-1 Combining Classifiers Trained on Subsets of Samples *(Polikar, 2006)*.

2.2.1 Heterogeneous Models

In heterogeneous ensemble methods, two or more different learning algorithms are applied on the same samples. For example, ANN, RBFN and Decision Tree algorithm could be applied

on the same task as base classifiers and their predicted outputs are then combined to from the final predictive of the classifier. Barry and Linoff shows that this approach is a good way of discovering meaningful rules and patterns (Berry & Linoff, 2011) in samples; in some cases it outperforms single or standalone predictive models. However, a number of studies show that the approach decreases the prediction error of the final model (Mishra, et al., 2010; Yeh, et al., 2012) to some extent.

With the error limitations the method has been applied in solving various classification tasks such as traffic incident management (Lee, et al., 2004), predicting and controlling crimes (Malathi & Baboo, 2011a; Malathi, et al., 2011b), predicting and optimising demand for Ambulance Services (Sasaki, et al., 2010) and emergency evacuation planning (Miah, 2011), banking and financial predictions (Jadhav & He, 2015), to detect medical fraud in health insurance (Lin & Yeh, 2012) and dynamic prediction of customer profitability over time (Chen, et al., 2019).

Despite the fact that several heterogeneous ensembles have been proposed their performance comparisons are difficult to evaluate because most of the methods are evaluated using different models and samples. Recent studies show that several challenges and problems are associated with heterogeneous models (Whalen & Pandey, 2013; Dunlavy & Gilpin, 2018). It also suffers from performance constraints (Day & Khoshgofaar, 2017). It also suffers from several other limitations such as merging the decision of different base classifiers (Tewari & Dwivedi, 2020) with different parametric training values. Another problem amongst others is the impact of their calibration on prediction accuracy of the model. This is a major challenge because of their high CPU usage and several parameters that are required to be optimized the base classifiers either in series or in parallel while training.

Even though few published studies have explored the ensemble heterogeneous predictive modelling, however many of these reports are purely descriptive and suffers from inadequate predictive output and are therefore inconclusive.

2.2.2 Homogeneous Models

On the other hand, ensemble homogeneous methods unlike the ensemble heterogeneous methods do not implement classifiers of their own. The concept however combines multiple weak classifiers of the same base classifier that are trained on the same distribution task to obtain the overall classifier output. The most common homogeneous models are the boosting, random forest and bagging techniques. However, each of the model has its own advantages and limitations. For example, Latinne, et al., 2001 argues that limiting the number of trees in random forests to minimum level could improve the prediction accuracy of the algorithm to some extent (Latinne, et al., 2001) before performance degradation occurs.

Recent evidence reveals that the ensemble size of base classifiers could significantly influence the prediction accuracy of ensemble models (HernáNdez-Lobato, et al., 2013). However there is a limited studies to address this concept. More recently a theoretical framework was proposed; "the law of diminishing returns in ensemble construction" (Bonab & Can, 2018; Bonab & Can, 2016). The framework shows that using a fixed number of independent base classifiers in ensemble models can perform more efficiently than using non-fixed number dependent of base classifiers.

Bootstrap Methods

Bootstrap is a meta-algorithm that is based on random resampling of datasets with replacement (Duda, et al., 2000). It uses statistics techniques to improve the accuracy of machine learning algorithms. The main advantage of bootstrap are mainly its ability to reduce

variance, a major problem in machine learning. However, review shows that there is a tenancy for overfitting and under-fitting issues that can lead to poor performance. This is because samples could appear more often than others and in some cases it does not appear at all while training the samples (Ghojogh & Crowley, 2019).

Bagging Methods

Bagging on the hand is an ensemble meta-learning (Breiman, 1996) algorithm that is based on bootstrap. Bagging constructs several independent base learner predictions on subsets of training samples with replacement. The concept uses bootstrap sampling to obtain the subsets of samples for training its base learners. Despite the models' ability to reduce variance, like bootstrap it could results in poor performance due overrepresentation or underrepresentation of training samples. Like the Bootstrap it has been demonstrated that this could lead to overtraining or undertraining of sample subsets. This is because during training subsets of sample data may appear more than once while others may not feature at all.

Stacked Generalization

This concept is based on using high-level models to combine lower-level models to reduce generalization error and achieve greater predictive accuracy (Wolpert, 1992). Stacked generalization or stacking is an ensemble technique that uses different learning algorithms algorithm to learn how to best combine the predictions from different models that are trained on the same dataset. The concept is graphically illustrated in Figure 2-2.

Stacking unlike the bagging and boosting methods, can be used as a hybrid-model to combine several homogeneous and heterogeneous algorithm types unlike boosting or bagging that

train learners of the same model. The outputs of the models are combined to compute the final prediction of any instance x such that:

$$\hat{y} = \sum_{j=1}^{m} \alpha_j \, h_j(x) \qquad 2.1$$

This is unlike boosting that sequentially computes weights α_j using an empirical formula. Stacking uses a level-1 algorithm that is meta-learner for learning the weights of the level-0 predictors. Such that the predictions of each instance x_j become the training data for the level-1 learner.

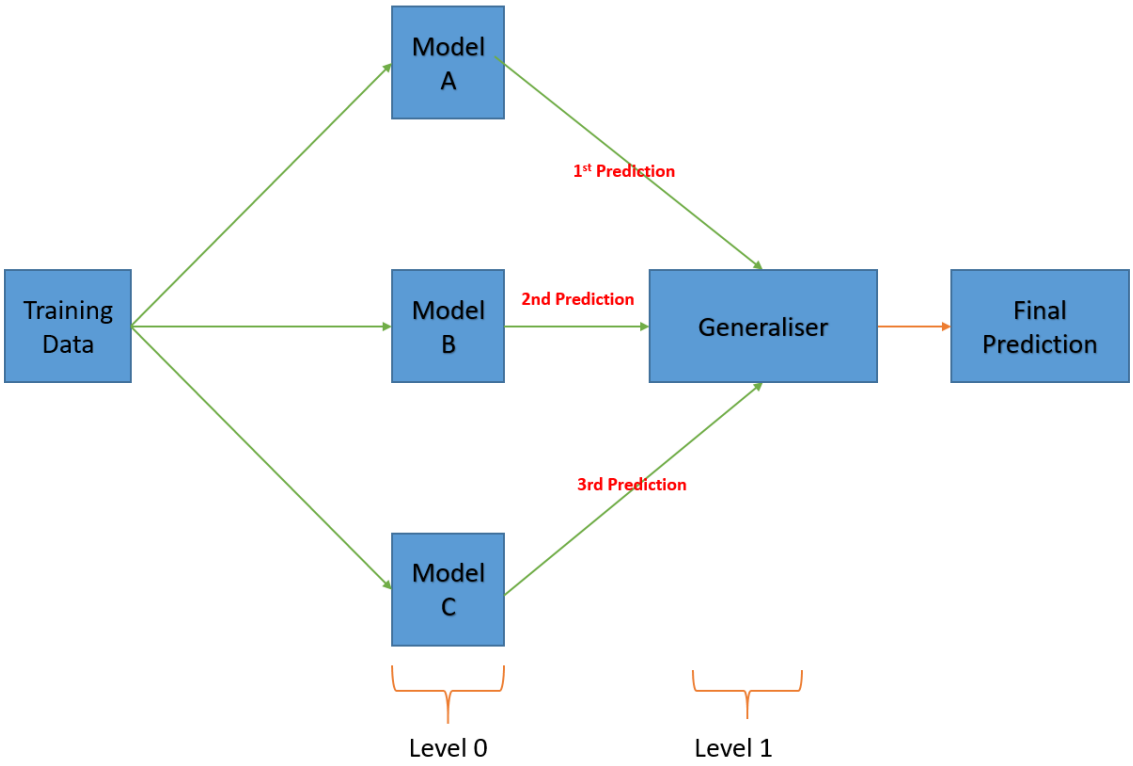


Figure 2-2: Architecture of Stacked Generalization [adapted from *(Freund & Schapire, 2014; Zhou, 2012)*]

It has been argued that theoretically, the concept of stacking can be viewed as a generalization of several ensemble methods (Zhou, 2012).

Previous study have also revealed that the method suffers from two main issues namely the required attributes at level-1 data and the type of level-1 learner to use in order to avoid generalization error (Ting & Witten, 1999); thereby improving performance of the concept. Therefore, the method is not widely used compare to bagging and boosting models.

2.2.3 Potentials and Drawbacks of Ensemble Models

Some of the factors that differentiate between the various ensembles methods (Rokach, 2005) among these are:

i. Inter-classifiers relationship - how classifier affects each other.
ii. Combining methods – the strategy for combining generated base classifiers together to form a single predictor.
iii. Diversity generator – since, diversity affect the performance of ensemble models, therefore to make the ensemble models efficient there should be a level of diversity between the generated hypotheses. Ensemble diversity is the differences in the decision or predictions made the ensemble leaners. Recent evidence suggests that combining only accurate learners is often worse than combining some accurate ones together with some relatively weak ones, since complementarity is more important than pure accuracy (Xu, et al., 2015). The performance of ensemble model is a balance or trade-off between the diversity of the classifiers and performance of individual classifiers.
iv. Ensemble size – the appropriate number of classifiers in an ensemble model.

Despite the diversification and aggregation ability of ensemble models over individual models; it also has a number of advantages and limitations. Some of these are discussed in this section.

2.2.4 Benefits and Potentials of Ensemble Models

Ensemble methods has the potential to trains several weak learners and then combine them to obtain a better final predictive output. The combination method of ensemble technique rather than randomly selection plays crucial role in the concept's ability to achieve strong generalization output compare to single classifiers. In his research into ensemble methods, Diettrerich attributed the good performance of ensemble methods to three main fundamental reasons namely statistical, computational and representational (Dietterich, 2000). Review also shows that the algorithm is easy to implement therefore make it an attractive learning technique.

However, ensemble techniques and most of the extended families in particular AdaBoost offers several theoretical, practical advantages and disadvantages (Schapire, 2013; Zhou, 2012) that limits their performances. However, it has been demonstrated that ensemble techniques improved predictive accuracy in several analytic and data mining applications (Rokach, 2010). Therefore it has been used in generating prediction hypothesis through collaborative learning (Arsov, et al., 2017) that are accurate to some levels. More recently, it has been used for predicting major cryptocurrency time series (Livieris & et, 2020) on hourly basis.

2.2.5 Problems and Limitations of Ensemble Models

Review shows that both heterogeneous and homogeneous methods suffer from several problems, however there is a lack of empirical performance comparison of the two methods. For example in heterogeneous method, there is a need to consider synchronization and communication between the base classifiers when evaluating which implementation method will give the best performance. Therefore, when one classifier finishes earlier than others that

are in synchronization, it must wait until other classifiers are ready. This can lead to wasted CPU, time wasting and diminished prediction performance accuracy (Wilson, 1996). Another problem with heterogeneous method is the issue of weak and strong collaborators amongst the base classifiers (Sublime, et al., 2015) which can also lead to misclassification problems if not accurately addressed or minimised.

Homogeneous classifiers on the other hand are sensitive to noisy data and outliers (Schapire, 2013), which can result in overfitting and subsequent poor classification accuracy. Review further shows that diversity among the ensemble weak models, decision making strategy, and the number of members to use (Kuncheva & Whitaker, 2003; Yang, 2011; Bian, 2007) to form the final classifier are challenges that affects the performance accuracy of boosting technique. The prediction performance of homogeneous classifiers also depends on the amount of training samples, type of base classifiers and the combination method of the samples. Several issues have also been associated with the algorithm (Wang, 2008; Galar, et al., 2012; Pintelas & Livieris, 2020) and its extended methods.

Like other machine learning concepts, difficulties arise in developing ensemble models. These include factors affecting the accuracy of ensemble models, class imbalance problems and to what level do they affect the predictive performance of the models (Wang, 2008; Sagi & Rockach, 2018). Other challenges among others include decision making strategy (He, et al., 2014) that ensures optimum inclusion, accuracy of individual models and diversity (Kuncheva & Whiitaker, 2003; Butler, et al., 2018) among the models.

A comparisons of existing ensemble models, problems and challenging in developing them are provided in Appendix A.1 Table A.1.

2.3 Loss and Cost Functions

The loss function is a mathematical method that uses the optimization concept (i.e. cost function) to enhance algorithm's performance as it learns in order to reduce task's predictive error. It accounts for the penalty paid for inaccuracy of predictions in classification problems. There are different types of Loss functions, however it can be classified according to the type of response of the prediction (Murphy, 2012; Natekin & Knoll, 2013) either as a classification or regression output.

For instance, AdaBoost is an optimization process that tries to fit an additive model based on a surrogate loss functions. Therefore, many variants of the algorithm have been developed based on different loss functions and other mathematical methods to address the issue of weight distributions (Zhou, 2012; Friedman, et al., 2000) among the weak learners in order to improve performance efficiency of the model.

Review shows that the choice of loss function is determined by the classification tasks. For instance, in a binary label problem loss function are used while in multi-class label task multi-class loss function are used. Therefore, the choice of loss function is primarily influenced by the characteristics of the dataset distribution and the classification objectives. Table 2.1 summarizes common loss functions and their derivatives that are associated with various ensemble methods.

Table 2.1 Common Loss Functions with their Derivatives [Adapted from *(Arora & Hazan, 2016)*]

Name	Formula (Loss)	Derivative	Algorithm
Squared error	$\frac{1}{2}(y_i - f(x_i))^2$	$y_i - f(x_i)$	L2Boosting
Absolute error	$\lvert y_i - f(x_i)\rvert$	$sign(y_i - f(x_i))$	Gradient boosting
Exponential error	$\exp(-\tilde{y}_i f(x_i)$	$-\tilde{y}_i \exp(-\tilde{y}_i f(x_i))$	AdaBoost
Logloss	$\log(1 + e^{-\tilde{y}_i f_i})$	$y_i - \pi_i$	LogitBoost

However, reviews further show that there are several problems that are associated with the loss functions, for instance, the exponential loss function puts a lot of weight on misclassified samples as the exponential curve shoot up on the left-hand side. This makes the function very sensitive to outliers and susceptible to misclassification problems. Therefore, loss function criterion used for classification should be able to penalize negative margins more heavily than the positives as the positive margins are already correctly classified (Hastie, et al., 2009). In general most studies have linked the performance of ensemble methods to crucial role that loss function plays in machine learning (Nie, et al., 2018; Hajiabadi, et al., 2020) and model optimizations. It is therefore essential to integrate a loss function that maximises output accuracy, minimises overfitting and generalization errors. The next section of the review focuses on the AdaBoost and its related variants.

2.4 Historical Timeline of Ensemble Models

Over the last decades, ensemble methods have been a major concern for researchers and practitioners in the area of machine learning and data mining. As a result, a significant number of several variants of the algorithm that addresses one problem or the other, in an attempt to enhance the performance of the method have been proposed. Many of which are instrumental to the development of several data science applications and contemporary

smart devices. This section provides a review of some of the historical development of ensemble variants in particular AdaBoost, their related theories for binary supervised learning problems. The review shows that several of the variants were based on theoretical and performance optimization properties of AdaBoost. Therefore, most of the variants were developed at targeting specific issues or problems that other variants could not tackle such as object detection, letter recognition, text categorization, class predictions, and optimization issues. The historical timeline of some of the common methods is illustrated in Figure 2-3. The arrows in Figure 2-3 indicates the direction of ensemble variants progress from year 1989 to year 2020. Firstly, the development of Boosting variant in year 1989 to the development of Modest AdaBoost/WaldBoost variant in year 2005. Secondly, the development of MildBoost/EmphasisBoost in year 2006 to the development of RegBoost variant in year 2020.

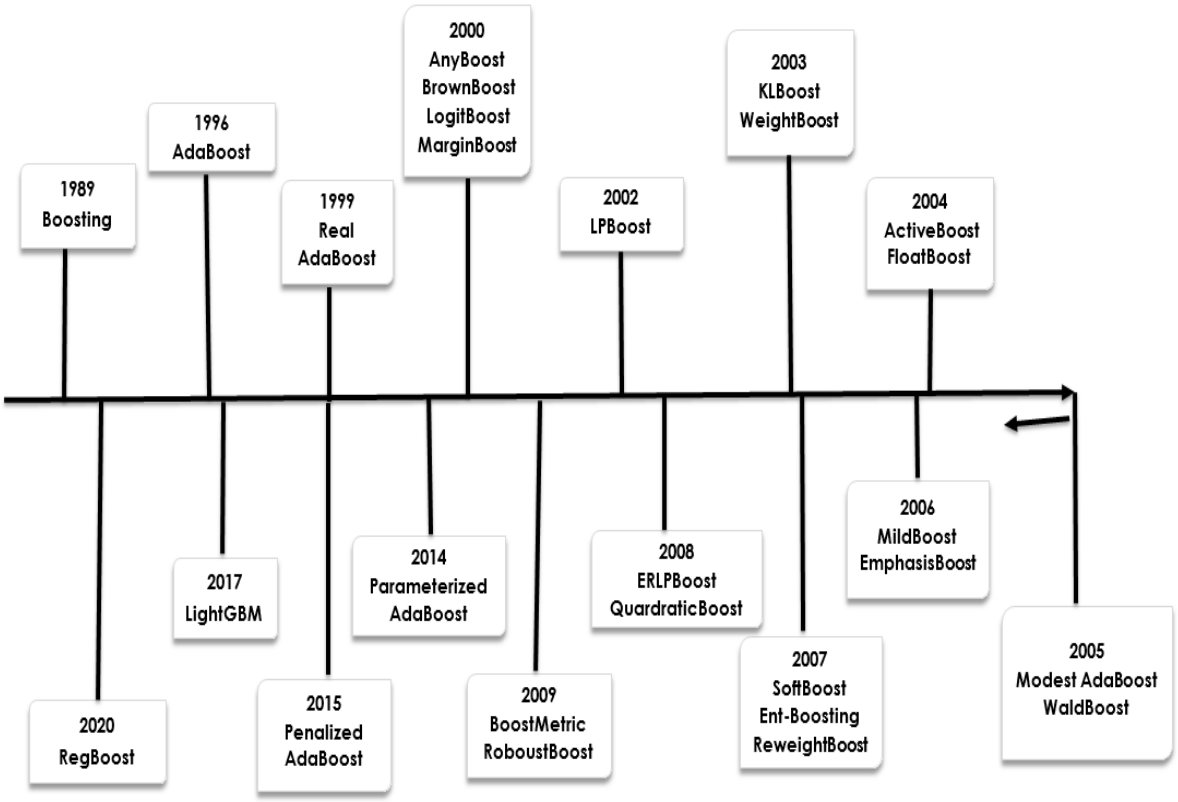


Figure 2-3 Timeline of Extended AdaBoost Variant [Adapted from Several Review Sources]

Boosting and Existing AdaBoost Methods

Boosting is an ensemble process of building strong classifier by combining sequence of several weak classifiers. It is an adaptive concept that maintains set of examples, class weights and focuses on base learners of the samples that are difficult to classify. Therefore, turns weak and difficult to train classifiers into strong learners. AdaBoost is an ensemble learning technique that is widely regarded as the most common family of boosting (Schapire & Freund, 2014) algorithms. It is also known as meta-learning and the name AdaBoost is coined from Adaptive Boosting. It was originally designed to boost the performance of binary clarifiers. It uses induction method to combine series of weak learners to form a stronger classifier instead of randomly selecting various classifiers. The predictions are then combined by a weighted majority vote for classification tasks and a weighted sum for regression tasks.

AdaBoost unlike other classifiers places extra weights on those cases that are most difficult to classify and therefore can be used to identify and remove outliers.

The method has been applied in many applications such as text classification, natural language processing, drug discovery and computational biology (Lin & Wang, 2011). Similarly in vision and object recognition (Viola & Jones, 2004), medical diagnosis (Thongkam, et al., 2008; Abuhasel, et al., 2015) identification, chemical fault detection and diagnosis (Karimi & Jazayeri-Rad, 2014) applications, however with numerous challenges and limitations.

In efforts to address some of AdaBoost limitations, numerous experimental studies has led to the introduction of variants of the algorithm by several authors. Albeit, many of these variants also have their own advantages, limitations and problems. For instance, Friedman proposed Gentle AdaBoost (Friedman, et al., 2000) in an attempt to improve the performance of Real AdaBoost variant. Contrary to the Real AdaBoost that applies the estimates of the weighted

class probabilities to perform training update, Gentle AdaBoost uses weighted least-squares regression during the training process. Though reviews show that Real AdaBoost converges faster compared to Gentle AdaBoost. However, Gentle AdaBoost is more stable compared to Real AdaBoost with respect to generalization error. Still, Gentle AdaBoost suffers from overfitting problems and increase in computation time especially when training noisy data. In addressing this problem another approach that limits the weight distortion based on a stretched distribution of the whole sample weights was introduced (Wu & Nagahashi, 2014), however with CPU challenges and other problems.

Another study to address some of the problems of Real AdaBoost and Gentle AdaBoost Modest AdaBoost algorithm (Vezhnevets & Vezhnevets, 2005) was proposed. The variant uses inverted weighing distribution scheme for the correctly and incorrectly classified weak classifiers and able to classify instances with small boundaries. Review shows that the algorithm outperforms Gentle AdaBoost in term of generalization error. However, the improved performance was compensated for by higher training errors also the performance evaluation of the algorithm are not stable (Wu & Nagahashi, 2015).

In an effort to address the generalization problem of ensemble AdaBoost, Real AdaBoost was introduced. However, the variant suffers from misclassification of samples that have been correctly classified. This limitation was further addressed by proposing Parameterized AdaBoost variant (Wu & Nagahashi, 2014) that uses parameter to penalize the misclassification of samples that are correctly classified. Experimental result of the variant shows that the variant improves the generalization error to some degree when compared with Real AdaBoost.

Penalized AdaBoost variant was also proposed (Wu & Nagahashi, 2015) in an attempt to address distortion problem in Gentle AdaBoost. The variant penalizes the misclassification of instances with small margins. It also introduces penalty policy that restrains the weight increase for instances with minimal margin. Experimental results show the method performs better that Gentle AdaBoost however with a similar training speed.

The unsatisfactory efficiency and scalability problems in Gradient Boosting Decision Tree (GBDT) when handling high feature dimensions large data is a major concern despite implementation of several optimizations methods. In resolving the problem, LightGBM (Ke & et.al., 2017) a variant that combines Gradient-based One-Side Sampling that resolves large number of data instances and Exclusive Feature Bundling that handles large number of features was proposed. Experimental results of the variant shows that it performs significantly better than XGBoost and GBDT in terms of computational speed and memory consumption.

Li *et*. *al*. in their work proposed the RegBoost variant (Li, et al., 2020), a multivariate regression ensemble algorithm based on the concepts of gradient boosted decision tree (GBDT), which combines the linear regression predictors from two distinct branches that are recursively trained to obtain the final predictive output. Though, the algorithm achieves similar performance as GBDT, however the method is only applicable to tasks with few datasets. Additional comparison review of ensemble AdaBoost variants and other related boosting methods are as presented in Appendix A.

This section provides an overview of some of the emerging ensemble methods with focus on their applications, problems and limitations. Ensemble based methods is valuable concept. It

improves the predictive output of ensemble based models compare with single or standalone classifiers. However, the issue hinders the performance of the technique.

Taxonomy Ensemble of Methods

Studies show that there is no standard taxonomy of ensemble learning. Therefore, there are inconsistencies of various methods and techniques that are currently proposed in the literature. For example, Witten and Frank (Anon., 2005) details four techniques of ensemble modelling; Bishop (Bishop, 2006) covers five methods; Marsland (Marsland, 2009) details three methods; Alpaydin (Alpaydin, 2010) covers seven methods. In their work Jain et al categorized taxonomy in ensemble methods into seventeen methods (Jain, et al., 2000) as shown in Table 2.2. Conversely, most algorithms can be categorized mainly in two ways. Firstly, grouping them by their learning styles either supervised (predictive) or unsupervised (non-predictive). Secondly, grouping them by their similarity in terms of their functions or how they work for example tree-based methods and neural network methods.

A number of studies suggests that a verse number of ensemble frameworks have been proposed, among these methods, the statistical approach has been the most intensively studied and used in practice (Jain, et al., 2000; Tsoumakas, et al., 2008). More significantly, review regarding importance of various predictive classifiers reveal that there is no objective conclusion about superiority of one classifier over the other. This is because the performance of any classifier depends on the nature of the problem, the type of dataset to be used and behaviour of the variables (Anjum., 2013).

Table 2.2 depicts the taxonomy of machine learning techniques based on their combination methods and characteristics (Jain, et al., 2000).

Table 2.2 Taxonomy of Ensemble Combination Methods and their Characteristics (Jain, et al., 2000)

Scheme	Architecture	Trainable	Adaptive	Comments
Voting	Parallel	No	No	Assumes independent classifiers
Sum, mean, median	Parallel	No	No	Assumes independent features
Product, min, max	Parallel	Yes	No	Considers error correlation
Adaptive weighting	Parallel	Yes	Yes	Explores local expertise
Stacking	Parallel	Yes	No	Good utilization of training data
Borda count	Parallel	Yes	No	Converts ranks into confidences
Logistic regression	Parallel	Yes	No	Converts ranks into confidences
Class set reduction	Parallel cascading	Yes/No	No	Efficient
Dempster - Shafer	Parallel	Yes	No	Fuses non-probabilistic confidences
Fuzzy integrals	Parallel	Yes	No	Fuses non-probabilistic confidences
Mixture of local experts (MLE)	Gated parallel	No	Yes	Explores local expertise; joint optimization
Hierarchical MLE	Gated parallel hierarchical	Yes	Yes	Same as MLE; hierarchical
Associative switch	Parallel	Yes	Yes	Same as MLE, but no joint optimization
Bagging	Parallel	Yes	No	Needs many comparable classifiers
Boosting	Parallel hierarchical	Yes	No	Improves margins; unlikely to over-train, sensitive to mislabels; needs many comparable classifiers
Random subspace	Parallel	Yes	No	Needs many comparable classifiers
Neural trees	Hierarchical	Yes	No	Handles large numbers of classes

Figure 2-4 illustrates the taxonomy of machine learning techniques (supervised and unsupervised) based on their learning styles (Patrick, 2017).

Figure 2-4: Taxonomy of Machine Learning Techniques Based on their Learning Style (Patrick, 2017).

2.5 Radial Basis Function Network (RBFN) and Extended Kalman Filter (EKF)

Despite the availability of RBFN and EKF concepts, there is no reliable evidence that the two techniques have been utilized in addressing ensemble predictive performance issues. In one hand, RBFN is an alternative to multi-layer perceptron neural network algorithm that maps nonlinear inputs to linear outputs. On the other hand, there is a need to train the network parameters in order to optimize its predictive performance. However, this can be carried out in many ways unlike multi-layer perceptron network (MLPN) that mainly uses back-propagation concept.

EKF is an optimization procedure (that linearizes the estimate of the current mean and current covariance that reduces errors in predicted estimates). Recent studies suggest that the filter has been demonstrated to be a good candidate to train RBFN (Wang, et al., 2006; Chernodub, 2014) to enhance its predictive output performance. The application of EKF amongst others in various industries include object tracking, state estimation, navigation systems, GPS (Niu & Hu, 2016; Yang, et al., 2016). Therefore, the filter has been considered as a standard algorithm (He, et al., 2018; Wan & Merwe, 2002) in the theory of nonlinear estimation and transition models.

Theoretically, the main purpose of Kalman Filter is to minimize the mean square error between the actual and estimated data (Attarian, et al., 2012; Dróżdż & Szabat, 2016; Ramadurai, et al., 2012) which can serve as a good base classifiers trainers in order to enhance their predictive performance.

Therefore, in recent years several researchers have used EKF to train Neural Network models (Shareef, et al., 2007; Krok, 2013) and least square fitting (Lacey, 2018) with encouraging results however with challenges. Recent studies reports that EKF has a number of issues (Kurban & Beşdok, 2009; Kamath, et al., 2011) such as the need to improve the convergence of the filter and parameters initialization setting amongst others.

However, it has been argued the filter's estimation performance could be improved by intelligently initializing the training process and accurately determining the tuning parameters (Guerci, et al., 1994).

Since the process of estimating training weights of ensemble models can be considered as a discrete-data linear filtering problem. Therefore, EKF can be applied to train RBFN parameters. This approach will help to optimize the network parameters and improve the performance of the network in particular when training samples are limited or samples have partly missing (Liu, et al., 2018; Bohler, et al., 2021) values. However, there is no considerable evidence in the literature to show that EKF and RBFN concepts have been jointly applied to model ensemble tasks in particular with AdaBoost to boost ensemble performance.

2.6 Knowledge Gaps and the Need for Extending Ensemble AdaBoost Method

As has been noted there has been numerous streams of different research on how to optimize and improve the performance of ensemble predictive methods. Several of these streams of

research make a significant contributions to the literature of ensemble predictive models and help to understand the performance factors and problems of several techniques (Anjum., 2013; Chou, et al., 2013; Freund & Schapire, 1977; Raskutti, et al., 2014). However, several studies show that current ensemble models suffer from numerous performance issues (Wang, et al., 2015; Cao, et al., 2012; Kim, et al., 2017; Xiao, et al., 2018; Merjildo & Ling, 2012) limitation and problems. A number of important limitations need to be considered amongst other are:

i. The need for large amount of historical data to train the base classifiers
ii. Application of complex empirical equations that are difficult to implement and may introduce prediction errors.
iii. Problem with base classifiers that are not capable of handling complex and imbalance samples.
iv. Overfitting and generalization problems.
v. Loss function and optimization difficulties
vi. Complexity and topology of ensemble models in terms of time and space (Jimenez, 1998; Gao & Xu, 2014)

As technology advances and dependency on ensemble related technique increases (for example IoT, smart devices, etc.) the performance improvement of ensemble predictive methods in order to ensure their reliability are therefore inevitable.

Subsequently, investigation on how to improve their reliability, avoid the complexity of the models and their predictive performance is increasingly important. Despite the fact that several models, theories and conceptual models have been proposed in the past, nevertheless none of these techniques have included and tested the integration of EKF, RBFN and AdaBoost as an ensemble technique.

2.7 Evaluation Criteria

For a two-class or multi-class task, the ultimate accuracy measurement is generally based on the overall performance of the algorithm. However, review shows that it is possible for a clarifier to achieve a very high accuracy in large classes due to bias but performs poorly when predicting from smaller or unbalanced classes. This can be misleading, therefore, it is important to consider other forms of performance measures to ascertain the performance and reliability of predictive algorithms. Classification performance of a two-class task can be represented by a 2x2 confusion matrix as depict in the Table 2.3

Table 2.3: Confusion Matrix for a Binary Classifier.

	Predicted Classes	
Actual classes	**Predicted - No**	**Predicted – Yes**
Actual - No	True Negative - TN	False Positive –FP
Actual - Yes	False Negative - FN	True Positive -TP

As illustrated in the table it produces four types of predictive outcomes on a two-class task:

i. True positives – These are the correctly classified positive cases that belongs to the positive class.

ii. True negatives - These are the correctly classified negative cases that belongs to the negative class.

iii. False positives - These are the incorrectly classified positive cases that actually belongs to the negative class.

iv. False negatives - These are the incorrectly classified negative cases that actually belongs to the positive class.

Single Class and Other Performance Measures

As previously highlighted, accuracy can be a misleading measurement in a model that involves large class of imbalance samples. It is possible for such model to predict majority class for all possible predictions with a high classification accuracy which does not reflect minority class predictions. To address this issue a number of standard metrics that are related to information retrieval area have been adopted to further evaluate the performance (Gunawardana & Shani, 2009) of a binary predictive classifier. For a binary class Recall or the Sensitivity (R) or the True positive Rate (TPR), the Precision (P), the F-Measure (F1-Score) and the Specificity (S) or the True negative rate (TNR) are defined as follows.

$$R = TP/TP + FN \quad (2.1)$$

$$P = TP/TP + FP \quad (2.2)$$

$$F = 2 * P * R/P + R \quad (2.3)$$

$$S = TN/TN + FP \quad (2.4)$$

Precision is a measure of how many of the correctly predicted cases actually turned out to be positive. Recall is a measure of how many of the actual positive cases were able to be predicted correctly with our model. On the other hand, the traditional F1-score is a harmonic mean of Precision and Recall. It gives a combined idea of the Recall and the Precision metrics and has been regarded as a better measure of the incorrectly classified cases than the accuracy metric.

2.7.1 ROC and AUC Curve Metrics

The receiver operating characteristic curve (ROC) is a graph showing the performance of a classification model at all classification thresholds as depict in Figure 2-5. The curve takes TPR and FPR as parameters on the y-axis and x-axis respectively. The area under curve (AUC) measures the entire 2-D area underneath the entire ROC curve. The AUC provides an aggregate measure of performance across entire likely classification thresholds (Mandrekar, 2010; Bradley, 1997).

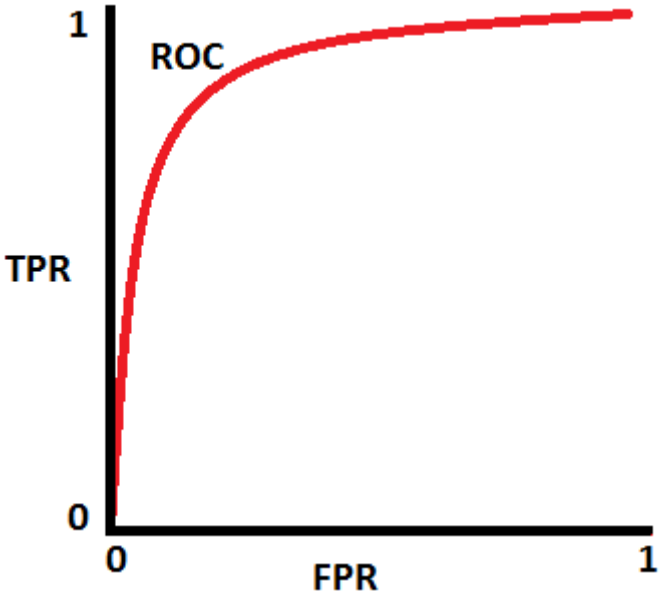


Figure 2-5: The ROC and the AUC (adapted from: (MBASkills., 2021)].

2.7.2 Other Evaluation Metrics

RMSE, MAE and R^2 are other statistical metrics that are adopted to assess the performance evaluation of classifiers. These are defined as:

$$MSE = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{2.4}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i) \tag{2.5}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \hat{y}_i)} \tag{2.6}$$

When the values of RMSE or MAE are small it shows the prediction is close to the expected value. On the other hand, when the value of R^2 are large or 1 at most, it shows that the

predicted value is close to the expected value. Additionally, other related equations and expressions are described in the subsequent chapters of the thesis.

2.8 Class Imbalance in Ensemble Modelling

Class imbalance occurs when the number of samples that represent one class is much lower or greater than the ones of the other classes of the same sample. Review shows that many of the well-established models are unable to adequately learn from imbalance distributions. Therefore, majority of the traditional models predicts poorly with skewed minority class distribution resulting in unsatisfied suboptimal results (Batuwita & Palade, 2013; Wang, et al., 2017) such that the model cannot be applied for future predictions. Negative and unpredictable effects of imbalance class classifiers have also been reported in the literature such as classification performance (Luque, et al., 2019; Vuttipittayamongkol, et al., 2021; Peng, et al., 2021) problems.

Today various methods have been developed and introduced to address the problem of class imbalance in ensemble modelling such as over-sampling method (Wang, et al., 2017), margin theory (Feng, et al., 2018), constraint projection (Guo, et al., 2020), cost-sensitive, data pre-processing and hybrids approach (Wu, et al., 2019).

Majority of the suggested techniques to address the problem are either at the data-level or at the algorithmic-level.

At the data-level are several approaches such as resampling methods and manipulation of the training data to correct the skewed imbalance class distribution such as data sampling.

At the algorithm-level are various methods such as cost-sensitive, and hybrid or ensemble (Leevy, et al., 2018; Blagus & Lusa, 2013) methods that ensures the accuracy of the imbalance class during sample training.

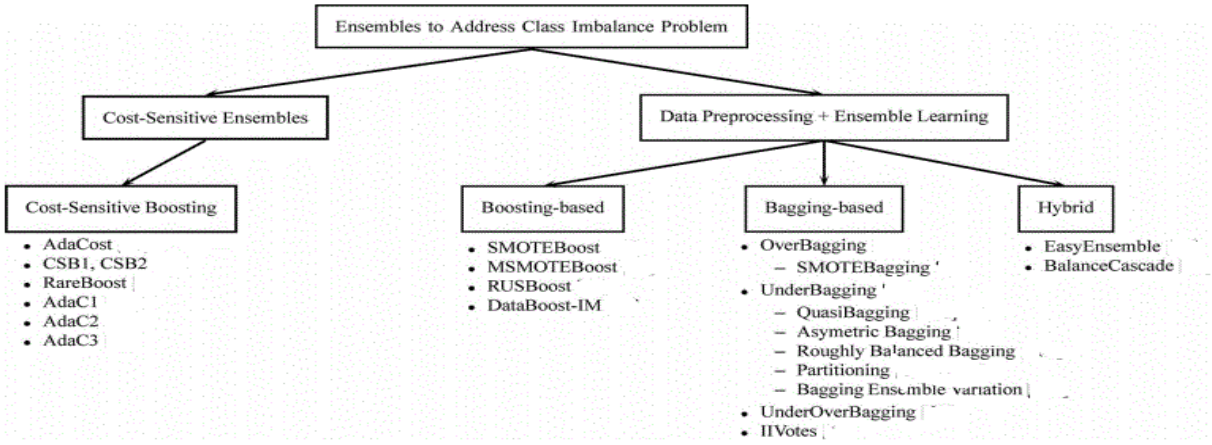


Figure 2-6: Taxonomy of Ensemble Techniques for Imbalance Problems *(Galar, et al., 2012)*.

Galar *et. al*. proposed taxonomy for ensemble methods that was based on techniques that are used to address the imbalanced datasets (Galar, et al., 2012). The taxonomy is as depict in Figure 2-6. The focus of the taxonomy is to address the class imbalance where each method can be categorized depending on the ensemble method that is used in addressing the class imbalance problem either at the data-level or at the algorithm-level. In this research several methods were considered and implemented by exploring resampling techniques, reweighting and other methods.

2.9 Diversity and Ensemble Classification

Diversity in ensemble model is the difference or disagreement among the generated base classifiers that are combined to form the final classifier. A number of authors have demonstrated that substantial diversity among classifiers play important factor in ensemble accuracy (Kuncheva & Whiitaker, 2003; Butler, et al., 2018) when combining the various weak

classifiers. However, review shows that determining diversity among the classifiers is not a straight fordward concept.

In regression estimators, Brown, et al. (2005) argued that diversity can be “formulated in terms of the covariance between individual estimator outputs, and the optimum level is expressed in terms of a bias-variance-covariance trade-off”. However, Reeve and Brown argued that to accomplish optimum performance in ensemble classifier the diversity among the weak classifiers and loss function in regression task is essential (Reeve & Brown, 2018). Similarly, in classification the important of diversity has also been emphasised (Mellor & Boukir, 2017) in order to obtain an optimum predictive performance.

In their work, Kuncheva and Whitaker examined the importance of diversity based on ten different statistical metrics which can measure diversity (Kuncheva & Whitaker, 2003) between different binary classifier outputs and their relationship with ensemble classification accuracy. However, several other studies show that there is a trade-off between ensemble classification accuracy and the diversity (Chandra, et al., 2006; Reeve & Brown, 2018) of the different ensemble classifiers to achieve an optimum accuracy. That is the more diverse the classifiers are the more accurate the classification output. This shows that there is no perfect classifier (Park & Cho, 2003; Guo, et al., 2013), therefore in machine learning the objective is to look for the best set of weak classifiers with different errors and prediction accuracies; then combine them with the best possible combination methods under appropriate conditions to obtain a better classifier.

2.9.1 Diversity Measures

There are different diversity measures many are available from the mainstreams of statistics and statistical pattern recognition that are primarily developed to address classifier ensembles (Kuncheva, 2014) problems.

In her work Kuncheva shows that quantitatively diversity can be defined in terms of pairwise and non-pairwise comparisons. The pairwise consider a pair of a classifier at a time: for example an ensemble L classifiers will produce $L(L-1)/2$ pairwise diversity the average across all possible pairs is taken as the as the diversity of the pool.

To illustrate diversity calculation, consider a table of the joined outputs of classifiers D_i and D_j as shown in Table 2.4. The entries in the table are the probabilities for the respective pair of correct or incorrect outputs.

Table 2.4: A 2 × 2 Relationship Table with Probabilities.

	$D_i$correct(1)	$D_j$wrong(0)
$D_i$correct(1)	a	b
$D_j$wrong(0)	c	d
Total, a + b +c + d = 1		

The disagreement measures, the probability that two classifiers h_i and h_j will disagree on their prediction, that is, two classifiers making different predictions (Kuncheva, 2014; Zhou, 2012; Ho, 1998), is illustrated in Equation 2.7.

$$Dis_{i,j} = b + c \tag{2.7}$$

Where $b\ and\ c$ are the probability values, as illustrated in Table 2.4.

The value of disagreement is in$[0, 1]$, the larger the value the larger the diversity. On the other hand, the non-pairwise access the entire ensemble diversity together rather than averaging pairwise measurements (Kuncheva, 2014; Zhou, 2012) and calculates directly one diversity for the ensembles. There are ten well established methods for measuring diversity (Kuncheva & Whiitaker, 2003) as summarized in Table 2.5.

Table 2.5: Summary of the 10 Diversity measures[1]

Name	**Notation**	**Direction**	**Pairwise/Non-pairwise**
Q-statistic	Q	↓	Pairwise
Correlation coefficient	P	↓	Pairwise
Disagreement measure	D	↑	Pairwise
Double-fault	DF	↓	Pairwise
Kohavi-Wolpert variance	KW	↑	Non-pairwise
Interrater agreement	K	↓	Non-pairwise
Entropy measure	Ent	↑	Non-pairwise
Difficulty Measure	Θ	↓	Non-pairwise
Generalized diversity	GD	↑	Non-pairwise
Coincident failure diversity	CFD	↑	Non-pairwise

[1] The arrow specifies whether diversity is greater if the measure is lower (↓). The higher the value the less diverse (i.e. similarity or greater (↑)) the classifier. The higher the value the more diverse (Kuncheva & Whiitaker, 2003) is the classifier.

2.9.2 Generalised Diversity

It has been demonstrated that the diversity of ensemble is maximized when the failure of one classifier is accompanied by the correct prediction of another classifier; minimum diversity occurs when a failure of one classifier is always accompanied by a failure of the other classifier (Kuncheva, 2014; Kuncheva & Whitaker, 2003). Then the probability of both classifiers failing is the same as the probability of randomly picked classifier failing, that is $p(1)$ using

$$p(1) = \sum_{i=1}^{L} \frac{i}{L} p_i \tag{2.8}$$

$$p(2) = \sum_{i=1}^{L} \frac{i(i-L)}{L(L-1)} p_i \tag{2.9}$$

The generalization diversity measure GD is expressed as:

$$GD = 1 - \frac{p(2)}{p(1)} \tag{2.10}$$

Where,

p_i denotes the probability of randomly chosen classifier failing on a randomly from pool L. GD varies between 0, minimum diversity when $p(2) = p(1)$ and 1, maximum diversity when $p(2) = 0$.

2.10 Ensemble Classifier Outputs

The choice of classifier fusion algorithm of ensemble classifier L depends on the accuracy and the diversity of the ensemble members. Kuncheva identified four classes of classifier outputs. These are the abstract type, rank type, measurement type and the oracle type (Kuncheva, 2014). The abstract type produces a class level that is of the most universal type. The measurement level produces a c-dimensional vector that outputs between 0 and 1, each classifier output spanning $[0, 1]^c$ and contains the most amount of information. The rank level is suitable for problems with a large number of classes. The oracle output is artificial, this is

because it can only be applied to a labelled dataset. For a given dataset Z, the classifier D_i produces an output vector y_i such that:

$$y_{ij} = \begin{cases} 1, & if\ D_i \text{ classifies } object\ z_i \text{ correctly,} \\ 0, & otherwise \end{cases} \quad 2.11$$

2.10.1 Combination Methods

As previously described instead of trying to find the best single classifier from a set of base learners, an ensemble method resort to combination of the base classifiers to achieve a strong generalization output. This concept is variously referred to by various names like: committee of learners, mixtures of learners, classifier ensemble, consensus theory or multiple classifier systems. The benefits of this concept have been attributed to three fundamental reasons namely statistical issues, computational issues, and representational issue (Dietterich, 2000).

Therefore, when developing or enhancing an ensemble algorithm it is essential that the issues are adequately addressed to avoid performance failures. It has been noted that a learning algorithm that suffers from statistical issues will fail due to high variance. A learning method that suffers from the computational issue can be described as having high computational variance; a learning process that suffers from the representational issue is generally said to have a high bias (Zhou, 2012; Dietterich, 2000). Therefore, a through combination of the variance and bias of learning algorithm is essential for good predictive performance. Figure 2-7 shows how a committee of classifier works and how the results are combined to form the final prediction.

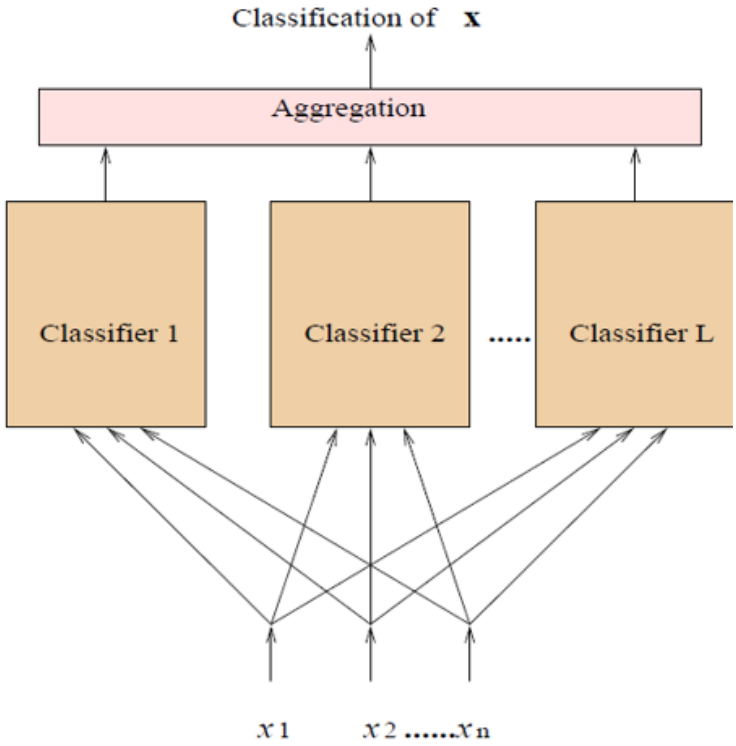


Figure 2-7: A Multiple Classifier Technique [Adapted from *(Polikar, 2006)*].

The feature value of object x, $(x_1, \ldots, x_n)$ are individually submitted to the L classifiers. The classifiers uses the features values of object x to predict an outputs of the features x inputs. The results from the classifiers are then combined to provide the final classification for object x that predicts the possible outcome. The main two approaches to this combination are:

Dynamic classification selection: A concept that predicts which of classifiers is most likely to be correct for each input object that can be assigned a single class label x.

Classifier fusion: This method assigns weight to all the case classifiers based on their importance in a multiple decision system and combine it to output the final classifier.

Kuncheva shows that technically, there are various aspects of ensemble classifier systems that one could alter or manipulate to improve the classification accuracy. These are the classifier

models, the feature subsets submitted to the classifier, the training sets or the combination of any of these (Kuncheva, 2014).

There are many experimental comparisons and methods in the field of statistical pattern recognition and machine learning to combine the outputs of the individual ensemble classifiers. Several well-established combination concepts have also been proposed, these among other are the majority vote, weighted majority vote, soft voting, naïve Bayes combination and multinomial methods (Kuncheva, 2014; Zhou, 2012). However, there are limitations and theoretical issues with some of these methods.

2.10 Early Stopping Models

Early stopping concept allows halting sample training once the model performance stops improving in order to avoid over training of the sample that can lead to overfitting. The technique of early stopping modelling methods are widely applied on process based models (Wu & Liu, 2009; Simsek & Turk, 2016; Wei, et al., 2019; Lauer & Bloch, 2005) to enhance their predictive performances with promising results. Albeit, the challenge is to train the network to a point such that it is capable of learning and adequately mapping the input objects to the corresponding output objects, that is to the point of eluding the problem of under-fitting or overfitting of the model while training. However, the irony about early stopping concept is that too many epochs can lead to overfitting of the training dataset, at the same time too few epochs can also result in under-fitting the model (Yu & Zhu, 2020; IBM, 2021; Anon., 2020).

2.10.1 Early Stopping Strategy

The strategic focus of early stopping model is on the learning speed in order to avoid overfitting (Ying, 2019; Botvinick, et al., 2019) and under fitting problems. The concept is one of the regularization techniques that is used to stop sample training as soon as the validation

loss starts to stray away from the training loss function. As illustrated in Figure 2-8, originally both training and validation loss start decreasing as the algorithm learns something about the data. Then at some point the validation loss starts to increase while the training loss continues to decrease, this is often when over-fitting starts to happen (ICL., 2021; Ying, 2019). Therefore, if the model continues learning after this point, the validation error will increase while the training error will continue decreasing. Likewise, if the model stop learning before this point, it will lead to under fitting and if the model stop learning after the point it will also result to over fitting.

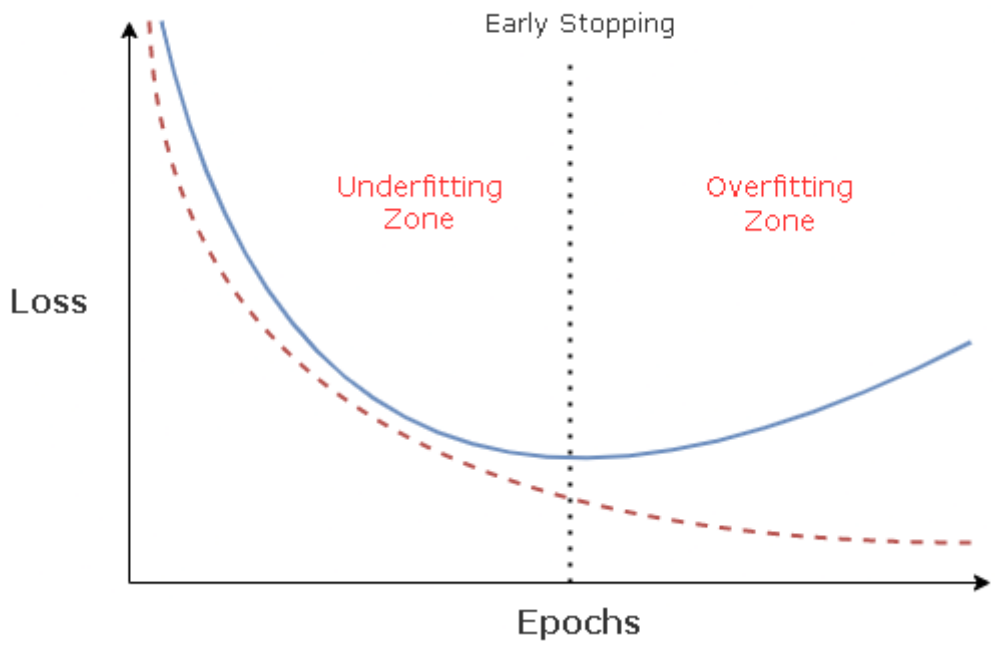


Figure 2-8: Overfitting Point and Early Stopping – Validation and Training [Adapted from *(Igareta, 2021)*].

Recent evidence suggests that one way model can over fit is when its weights are allowed to grow out of control (ICL., 2021; Nielsen, 2015). In the case of artificial neural nets, the learning process is to find a perfect set of weights and bias. The neurons learn at a rate that is determined by the partial derivatives of the cost-function, $\frac{\partial c}{\partial w}$ and $\frac{\partial c}{\partial b}$. It has been

demonstrated that the speed of learning depends on the values of the two partial derivatives, such that:

$$\frac{\partial c}{\partial w_j} = \frac{1}{n}\sum_x X_j(\sigma(z-y) \tag{2.12}$$

$$\frac{\partial c}{\partial b} = \frac{1}{n}\sum_x (\sigma(z-y) \tag{2.13}$$

Where, w_j is the j^{th} weight, b is the bias, C is the cost, X_j is the j^{th} input and y is the output.

2.10.2 L1 and L2 Regularization Methods

It has also been demonstrated that over-fitting problem can be penalized by applying L1 and L2 regularisation (ICL., 2021; Karim, 2018; Valverde, 2018). The notions of L1 and its derivatives are as illustrated in Equation (2.14) and Equation (2.15) respectively. Likewise the notions of L2 and its derivatives are as illustrated in Equation (2.16) and Equation (2.17) respectively.

The λ indicates the regularization size.

$$J(\theta) = Loss(y, \hat{y}) + \lambda \sum_w |w| \tag{2.14}$$

$$w \leftarrow w - \alpha(\frac{\partial Loss}{\partial w} + \lambda sign(w)) \tag{2.15}$$

$$J(\theta) = Loss(y, \hat{y}) + \lambda \sum_w w^2 \tag{2.16}$$

$$w \leftarrow w - \alpha(\frac{\partial Loss}{\partial w} + 2\lambda w) \tag{2.17}$$

In one hand L1 regularization tries to estimate the median of the data while on the other hand the L2 regularization tries to estimate the mean of the data to avoid overfitting the model.

2.11 Chapter Summary

This section is a discussion on the background knowledge and reviews of relevant literature on predictive ensemble methods. It includes deliberations on previous works, problems and limitations in order to identify knowledge gaps.

Amongst these problems, challenges and limitations are:

i. Ensemble method require large historical data and is difficult to obtain and to train.
ii. Ensemble methods require complex empirical equations. These are difficult to implement and may introduce prediction errors.
iii. The issue of topology and complexity of ensemble methods in terms of space and time (Gao & Xu, 2014).
iv. The generalization and optimization difficulties of the ensemble techniques.

There are various concepts and methods that have not been considered in the literature that can be applied in addressing some of these problems and limitations. Some of these approaches amongst others include:

i. The integration of EKF, RBFN and AdaBoost concepts as an ensemble technique to address ensemble issues.
ii. Statistical approach of threshold separation and stopping sample training early to avoid overfitting problems
iii. Application of identified breast cancer prognostic features to develop a multivariate ensemble logistic regression model based on features of the sample.

iv. The development of non-invasive breast cancer survivability visual tool that integrates theoretical concepts of ensemble research and practice as demonstrated in this thesis.

Section one reviews ensemble learning methods and their limitations. Section two reviews existing boosting methods from which the main research gaps are identified. Section three reviews loss function optimization methods and diversity in ensemble methods. Section four reviews Radial Basis Function Network (RBFN) and Extended Kalman Filter (EKF) and section five discussed and identified research gaps. Section six and seven discusses ensemble evaluation criteria and class imbalance. Sections eight and nine presents diversity and ensemble classification, and ensemble classifier output respectively. Section ten provides brief description of regularization methods. Finally, section eleven provides summary of the chapter. In the next chapter, the research methodology and the philosophy underlying the epistemological and ontological of the research will be discussed.

Chapter 3: Boosting Methods, Standalone and Ensemble Models

3.1 Introduction

The previous chapter discussed relevant models and theories proposed in the past to enhance the performance of ensemble models. In this chapter and subsequent three chapters, the aim is to provide an overview of the research procedures exploited in this study that leads to the selection of appropriate research methods for guiding the proposed ensemble models and their performance validations in this thesis. However, due to the nature of the research tasks, mixed research methods were implemented to enable the accomplishment of the aims and objectives of the research highlighted in the previous chapter. As presented in the previous chapter, the study was divided into four phases. This chapter covers Phase One of the study.

The focus of this phase is to address Objective 1 of the work. To achieve this objective several empirical investigations on boosting methods and related theories were carried out using different predictive techniques. This was realised by conducting various experimental studies, simulations and modelling of ensemble AdaBoost, standalone algorithms, and ensemble methods in order to understand, and establish some of the features that affect the performance of standalone and ensemble methods. This was followed by the analysis and discussion of the various results.

3.2 AdaBoost and Boosting Methods

Boosting is an ensemble meta-algorithm for primarily reducing bias, and variance in supervised learning algorithms. It is a family of boosting machine learning algorithms that convert weak learners to strong learners. Weak learners are model that are slightly better than random guessing, while strong learners are models that have good predictive output

that is well-correlated with the true classification (Ambati, 2021; Freund & Schapire, 2014; Zhou, 2012)

In general, AdaBoost algorithm can be viewed as a Neural Network algorithm in which the weak and the strong learners form the hidden part of the network for the hidden and the output parts of the network respectively as illustrated in Figure 3-1. Based on the concept of AdaBoost strategies; using several Neural Network (NN) algorithms as base classifiers while training the samples. These committees of NN base classifiers are then combined to obtain the final predictive output via a combination function, $F(.)$ to obtain the final predictive output y.

As presented in Chapter 2, AdaBoost has two main approaches of implementation. These are boosting by sampling and boosting by reweighting respectively. It is unclear from the literature review which of the two methods performs better than the other. Therefore it is crucial to carry out empirical studies to find out which of the methods actually performs better. Albeit, the difference between the two methods is a matter of their implementation and training techniques.

In reweighting method, AdaBoost minimizes the training errors by varying weights assigned to the samples during training based on their performance. Boosting by resampling method does not rely on training errors by varying weights, instead it uses the subset of training data in training its base classifiers which are then combined to form the final classifier.

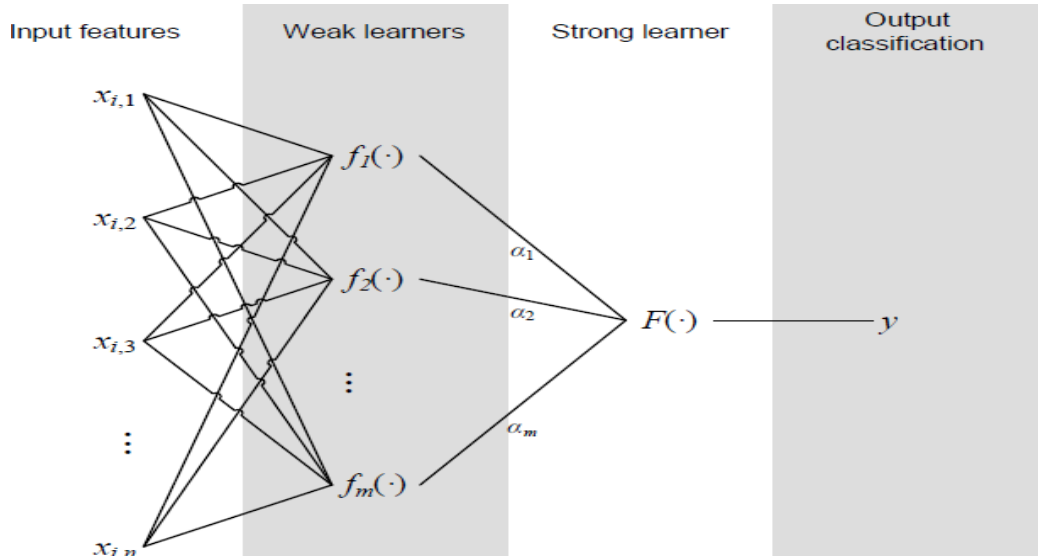


Figure 3-1: AdaBoost as an ensemble of Neural Network algorithm with activation functions [Adapted from *(Lee, et al., 2018)*].

3.2.1 Boosting by Reweighting

In boosting by reweighting the numerical weights for each sample are passed directly to the base learner. The base learner uses the weighting information when forming its hypothesis. Therefore, in boosting by reweighing equal distribution weights are assigned to the base classifiers at the beginning of the training. As training progresses the weight assigned to a classifier is increased if the classification is wrong. The weight is reduced if the classification is correct. Therefore, the weighted training error is minimized explicitly (Freund & Schapire, 2014; Zhou, 2012) to the exponential loss function.

3.2.2 Boosting by Resampling

Classically, boosting by resampling is also known as bootstrapping. It uses a given distribution Dt to generate an unweighted training samples by randomly selecting a sequence of examples S according to Dt, with replacement such that:

$$S = \left(x_{i_1}, y_{i_1}\right), \ldots, \left(\mathrm{x_{i_m}}, , \mathrm{y_{i_m}}\right) \qquad (3.1)$$

where each i_t is selected independently at random with replacement using the distribution Dt, from the original dataset which is fed to the base learning algorithm such that $\mathrm{x_{i1}}$ are the selected features and $\mathrm{y_{i1}}$ are the expected outputs of the selected features. The method consists of drawing pools of repeated samples from the original data population i.e. the selection of randomized repetitive cases with replacement. It therefore generates a unique sub distribution (sampling) that is based on the actual data population. Therefore, in resampling method there is a tendency that some examples will be included more than once in a given distribution while others might be omitted entirely. This can therefore, in one hand lead to overfitting of some samples when they are overrepresented. On the other hand it can also result in under fitting when some samples are underrepresented.

3.3 AdaBoost and Boosting Based Classification

3.3.1 Analysis of AdaBoost

Unlike most algorithms, AdaBoost automatically adapts to the error rate of their base classifiers and dynamically regulate the weight of the samples during training to enhance their predictive performance (Freund & Schapire, 2014). The combined output of the base classifiers is stronger and better than the individual classifiers acting alone on their own. However, review shows that the algorithm is susceptible to noise, and its accuracy also depends on the predictive accuracy performance of the chosen base classifier and tasks at hand (Zhou, 2012; Randhawa, et al., 2018). Figure 3-2 illustrates the basic flow chart for AdaBoost algorithm based on the resampling implementation method. The accompanying pseudocode is as illustrated overleaf.

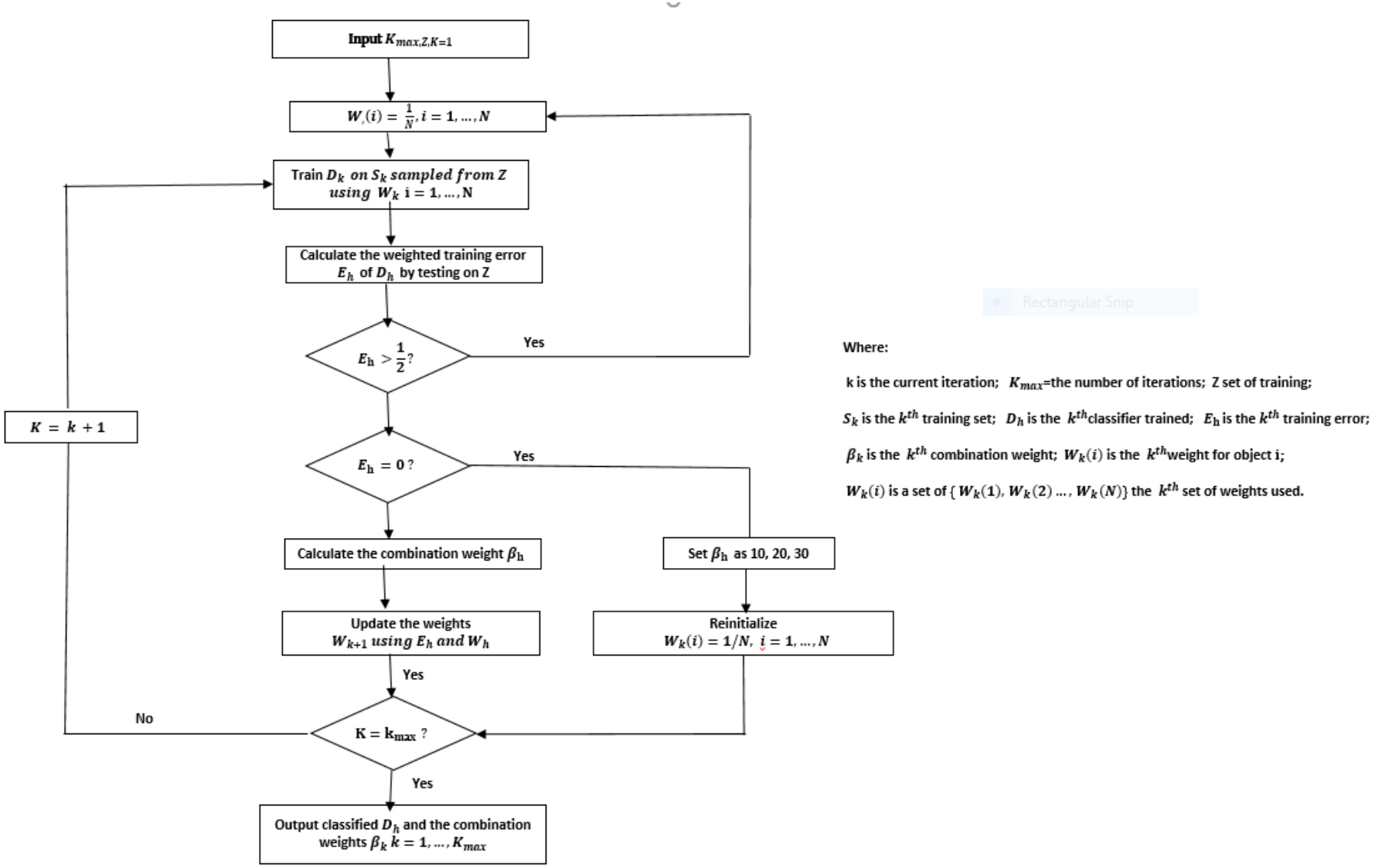


Figure 3-2: AdaBoost Algorithm Based on Resampling Implementation Method [Adapted from *(Freund & Schapire, 2014; Freund & Schapire, 1997)*]

In essence the final classifier $H(x)$ is computed as a weighted majority of the weak hypothesis h_t by vote where each hypothesis is assigned a weight α_t. This is given in equation (3.1) below. This is further demonstrated in Figure 3-2 above and the associated AdaBoost algorithm pseudocode below.

1. Given the training data : $\{(x_1,\ y_1), (x_2,\ y_2), \dots, (x_n,\ y_n)\} where\ x_i\ \in X, y\ \in Y = (-1, +1)$
2. Initialize the weight vector: $D_1\ (i) = 1/n$
3. For $t = 1, \dots, T$:
   i. Training a Weak Classifier using distribution D_t
   ii. Get weak classifier $h_t : X\ \rightarrow \{-1, +1\}\ s.t\ the\ error$
   iii. Calculate the error: $\varepsilon_t = Pr_{i\sim D_t}[h_t(x_t)] \neq\ y_i$
   iv. Compute the weight $h_t(x_t)$: $\alpha_t = 1/2 \log((1 - \varepsilon_t) / \varepsilon_t)$
   v. Update the weights:

$$D_{t+1}\ (i) = \frac{D_t\ (i)}{Z_t}\ X \begin{cases} e^{-\alpha_t}, & if\ h_t(x_t) = h_t \\ e^{\alpha_t}, & if\ h_t(x_t)\ \neq\ h_t \end{cases}$$
$$= \frac{D_t\ (i) \exp(\alpha_t y_t\ h_t(x_i))}{Z_t}$$

4. Out the final classification:

$$H(x) = sign(\sum_{t=1}^{T} \alpha_t h_t\ (x))$$

Pseudocode of AdaBoost Algorithm *(Schapire & Freund, 2014)*

$$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right) \tag{3.1}$$

The accuracy of the hypothesis is calculated as an error measure as illustrated in equation 3.2.

$$\varepsilon_t = pr_i{\sim}D_t[h_t(i) \neq\ y_i] \tag{3.2}$$

The weight of the hypothesis is a linear combination of all the hypotheses participating as shown in equation (3.3).

$$\alpha_t = \frac{1}{2}\ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right) \tag{3.3}$$

The distribution vector D_t is expressed as function many parameters illustrated in equation (3.4), where $\mathrm{Z_t}$ is a normalization factor such that the weights add up to 1 and makes $\mathrm{D_{t+1}}$ a normal distribution.

$$D_{t+1}(i) = \frac{D_t(i)\exp\left(-\alpha_t y_i h_t(x_i)\right)}{Z_t} \tag{3.4}$$

Theorem: Suppose the weak learning algorithm, WeakLearn when called by AdaBoost, generates hypotheses with errors $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T$. Freund and Schapire (1997) proved that the $error\ \varepsilon_t = Pr_i{\sim}D_t[h_t(i) \neq y_i]$ of the final hypothesis h_t output by AdaBoost is bounded by equation (3.6).

$$\epsilon = \mathrm{Pr_{i\sim D}}\left[\mathrm{h_f(x_i)} \neq \mathrm{y_i}\right] 2^{\mathrm{T}} \prod_{\mathrm{t=1}}^{\mathrm{T}} \sqrt{\epsilon_\mathrm{t}(1-\epsilon_\mathrm{t})} \leq \mathrm{e}^{-2} \sum_{\mathrm{t=1}}^{\mathrm{T}} \gamma^2{}_\mathrm{t} \tag{3.5}$$

where $\gamma_t = 0.5 - \varepsilon_t$. This shows that AdaBoost reduces the error at an exponential rate. It can be shown that to achieve an error $\varepsilon_.$, the number of learning round T is upper bounded by the natural logarithm (*ln)* of ∈ (Freund & Schapire, 2014; Zhou, 2012) as shown equation (3.6).

$$T \leq [\frac{1}{2\gamma^2} ln \frac{1}{\epsilon}] \tag{3.6}$$

Since the estimates can only be carried out on training data D such that $\in_D =$ $Pr_i \sim D_t[h_t(i) \neq y_i]$ i.e. the generalization error. Thus, the errors are training errors, while the generalization error $\in_D$ is more important. It has been demonstrated that the generalization error ($\in_D$) of AdaBoost (Freund & Schapire, 1997; Freund & Schapire, 1999) is upper bounded by equation (3.7).

$$\in_D \leq \in_D + \hat{O}\left(\sqrt{\frac{dT}{m}}\right) \tag{3.7}$$

The probability is at least $1 - \delta$ ($\delta > 0$), where d is the VC-dimension of the base learners, m is the number of training instances, T is the number of learning rounds and $\hat{O}(.)$ is the logarithmic terms and constant factors (Freund & Schapire, 1977; Freund & Schapire, 2014).

Where, $\in_D$= generalization error, T= number of learning round, m = number of training samples, $\delta > 0$ and $\lambda_t = 0.5 - \in_t$

3.3.2 Margin Explanations of Booting

The margin theory provides explanations to the success of AdaBoost as the key for characterizing its performance. Though Gao and Zhou(2003) and other researchers argue that AdaBoost does not usually over fit because it enlarge its margin after the training error reaches zero (Gao & Zhou, 2013) following a large number of training rounds. However, review further shows that this is a contradiction with the Occam's razor which prefers simple hypotheses to complex ones when both fit empirical observations well.

However, the generalization equation (3.9), and previous equations show that to achieve a good generalization. Therefore, it is important to constrain the complexity of the base classifiers and the number of learning rounds in order to avoid overfitting of AdaBoost (Zhou, 2012). In addition to this, Schapire *et. al.* (1998) proposed a margin-based theory that

provides margin explanation to AdaBoost. The authors argued that the theory is related to the distribution of margins of the training samples and the normalized margin of an ensemble as expressed in equation (3.9).

$$f(x)H(x) = \frac{\sum_{t=1}^{T} \alpha_t f(x) h_t(x)}{\sum_{t=1}^{T} \alpha_t} \tag{3.9}$$

Where α_t are the weights of the base learners. Based on the concept of margin the authors proved that, given any threshold $\theta > 0$ over the training sample D, with probability $1 - \sigma$, the generalization error of the ensemble $\in_D = P_{x \sim D}(f(x) \neq H(x)$ are bounded as

$$\epsilon_{\mathcal{D}} \leq P_{\boldsymbol{x} \sim D}(f(\boldsymbol{x})H(\boldsymbol{x}) \leq \theta) + \tilde{O}\left(\sqrt{\frac{d}{m\theta^2} + \ln\frac{1}{\delta}}\right) \tag{3.10}$$

$$\leq 2^T \prod_{t=1}^{T} \sqrt{\epsilon_t^{1-\theta}(1-\epsilon_t)^{1+\theta}} + \tilde{O}\left(\sqrt{\frac{d}{m\theta^2} + \ln\frac{1}{\delta}}\right) \tag{3.11}$$

where d, m, T and $\hat{O}(.)$ are as previously defined. Observation shows that above equations depend on the smallest margin as the probability $P_{x \sim D}(f(x)H(x) \leq \theta$. In addressing this problem Breiman proposed a variant of boosting algorithm $arc - gv$ that generates a larger minimum margins but with worse error compare with AdaBoost (Breiman, 1996)

$$\varrho = \min_{\boldsymbol{x} \in D} f(\boldsymbol{x})H(\boldsymbol{x}) \tag{3.12}$$

Such that in each round, $arc - gv$ updates α_t as

$$\alpha_t = \frac{1}{2}\ln\left(\frac{1+\gamma_t}{1-\gamma_t}\right) - \frac{1}{2}\ln\left(\frac{1+\varrho_t}{1-\varrho_t}\right) \tag{3.13}$$

Where γ_t is the edge of h_t and α_t is the minimum margin of the combined classifier up to the current round.

3.3.3 Statistical View of Boosting

A number of authors have shown the margin theory has a strong correlation with boosting and the support-vector machines (Vapnik, 1995; Awad & Khanna, 2015).

Unlike the other non-ensemble learning algorithms, after finding good weak hypotheses to combine in order to obtain good predictive rate while avoiding overfitting problem, the concept is interested in choosing a good coefficients α_t. Freund and Schapire argues that one realistic approach is to choose the coefficients such that the bound in equation 3.14 is minimized (Freund & Schapire, 1999).

$$\hat{\Pr}\left[\text{margin}(x, y) \leq \theta\right] + \bar{O}\left(\sqrt{\frac{d}{m\theta^2}}\right) \tag{3.14}$$

Similarly, in their work Friedman *et. al.* also argue that boosting can be viewed in terms of statistical principles of additive modelling and maximum likelihood (Friedman, et al., 2000). However the concept required modifications in order to optimize the process. Likewise, Schapire & Singer presents the exponential loss function of AdaBoost as a differentiable upper bound of the 0/1-loss function, a notion that is normally used for measuring misclassification error (Schapire & Singer, 1999).

Therefore, different explanations and various ambiguity have led to the development of variants of the algorithms by considering different surrogate loss functions and related methods to improve predictive performance of the algorithm as presented in the previous chapter.

It has been demonstrated that when regarding boosting procedure as an optimization of a loss function, then a mathematical programming can be applied to maximize the weights of weak learners (Demiriz, et al., 2002; Warmuth, et al., 2008; Xie, et al., 2019) and therefore

enhance its performance. This approach is adopted in subsequent sections of this report where various extensions of the ensemble models are established and demonstrated in improving the predictive performance of algorithm.

Therefore, considering an additive model $\sum_{h \in H} \alpha_h \, h$ of pool H of committee of weak learners, where μ_i is the loss of the model on instance x_i. It has been demonstrated that the sum of the coefficients and losses (Demiriz, et al., 2002) is bounded as illustrated in Equation 3.15.

$$\sum_{h \in H} \alpha_h \; + \sum_{i=1}^{m} \mu_i \leq B \tag{3.15}$$

According to Zhou, equation (3.15) bounds the complexity (Zhou, 2012) of the model such that the generalization error is bounded by equation (3.16) (Zhang, 1999).

$$\in_D \leq \; \hat{O}\left(\frac{\ln m}{m} B^2 \ln(Bm) + \frac{1}{m} ln \frac{1}{\partial}\right) \tag{3.16}$$

Where $B \geq 1$ and $\alpha_h \geq 0$.

Therefore, considering T weak learners such that $y_i = f(x_i)$ *is* the label of training instance x_i, and $H_{i,j} = \; h_j(x_i)$ is the output of weak learner h_j on x, such that the optimization task is as illustrated in equation (3.17).

$$\min_{\alpha_j \mu_i} \sum_{j=1}^{T} \alpha_j + B \sum_{i=1}^{m} \mu_i \tag{3.17}$$

3.4 Empirical Investigation on Algorithmic Methods.

3.4.1 Experimental Investigation on Boosting Implementation Methods

One of the main interests in this study is to compare the accuracy performances of some of the commonly used ensemble implementation methods based on different base classifiers and datasets. Firstly, it will assist in examining the performance and correlation between the

outputs of the two implementation methods. Secondly, help to determine which method is easier to implement, how it can be improved, how they relate to each other under different experimental sources. In addition, it will provide answers to Hypothesis 1 and Hypothesis 2 of the thesis.

3.4.2 Base Classifiers

In this section a brief description of the algorithms that are used during the experimental investigation to train and build the classifiers that map the input data to output category or class are provided. The algorithms are: Decision Stumps, Neural Network, Random Forest, Support Vector Machine (SVM) and AdaBoost as a meta-algorithm. The algorithms were selected based on their popularity and application in diverse projects (Schapire & Freund, 2014; Witten, et al., 2011) and ability to handle training by reweighting and resampling methods, and the classifiers' cost matrix compatibility with the training datasets.

Decision Stumps

Boosting with decision stumps is quite popular in data mining and have been shown to achieve better performance compared to unbounded decision trees due to their simplicity. It consists of a one-level decision tree and one internal node which are connected to the terminal nodes. It makes prediction based on the value of a single input feature (Witten, et al., 2011). Therefore, it has been used as a base classifier in various machine learning models such as Viola-Jones' face detection framework that employs AdaBoost as a main classifier (Viola & Jones, 2004).

Artificial Neural Network

The artificial neural network (ANN) is a family of artificial intelligence models like the biological brain that is capable of mapping complex and non-linear functions that depend on many inputs into linear outputs. It is generally presented as systems of interconnected neurons that exchange messages between each other. Like AdaBoost, ANN connections have numeric weights that can be changed thereby making neural networks adaptive to inputs and capable of learning and solving complex problems (Witten, et al., 2011). However, in this study we used the multilayer perception (MLP) with back-propagation which is a popular architecture of ANN (Haykin, 2008). During the study, we observed that ANN models took a longer time to run compared with other models. This is due to the complexity of the algorithm's architecture and topology such as the number of neurons, number hidden layers and other parameter settings.

Random Forest

Random forest is an ensemble of several decision trees. It is a meta-estimator algorithm that fits several decision tree classifiers on various sub-samples of the dataset. It operates by constructing many decision trees at training time. It then outputs the class that is the mode of the classes as mean prediction of the individual trees.

Support Vector Machine

SVM has its roots in statistical learning theory and represents the decision boundary using a subset of the training examples known as support vector (Cheng & Tan, 2010; Cheng, et al., 2007). It is a discriminative classifier that is formally defined by separating hyperplanes and has been successfully used in many classification and regression analysis problems. In Support Vector Machine new examples of data are mapped into points in a separating

hyperplane and are predicted to belong to a category based on which side of the hyperplane the data falls on (Witten, et al., 2011; Corinna & Vladimir, 1995).

Recent studies show that the algorithm works well with high-dimensional data and avoids the dimensionality problem that other algorithms faces.

3.4.2 Experimental Investigations: Ensemble Methods

The focus of this experimental study is to understand the performance of boosting methods. To see whether there is a link between boosting by reweighing, boosting by resampling and accuracy of ensembles modelled by using AdaBoost algorithm. To investigate if there is a connection between ensemble accuracy and complexity of the base classifiers. For the experiment a total of 13 different datasets obtained from the Machine Learning Repository (Dua & Karra Taniskidou, 2017) as illustrated in Table 5 were used. The reason for doing this is to ensure that the datasets used in the case study are derived from different domains to generalize outcomes of the study.

Ten-fold cross validation approach (i.e. splitting the original datasets into K (10) folds or subsets of equal size) was applied to estimate the performance of the classifiers (Delen, et al., 2005) as illustrated in Figure 3-3. This method has the advantage of using all the datasets for both training and testing the model thereby avoids overfitting problems. The stratified cross-validation tend to generate comparison results with lower bias and lower variance.

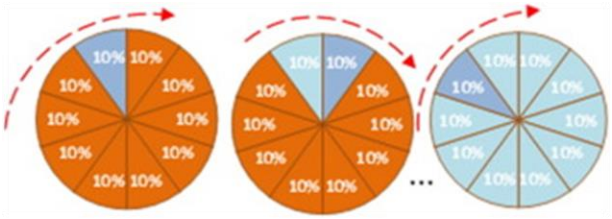


Figure 3-3: Illustration of the 10-Fold Cross Validation Technique *(Delen, et al., 2005)*.

The datasets were divided into learning and validation samples, the learning samples were used for acquiring rules. The validation samples were used for validating the acquired rules obtained during the training. The process is carried out ten times.

i. For each fold train the classifier using all the folds except one (K-1)
ii. Use the left out fold to test the model by calculating the cross validation metrics
iii. Run the k-fold cross validation run several k times (that is 10 times)
iv. Average the validation metrics across the subsets to get the final cross validation metrics.

The following learning algorithms were implemented as base classifiers during the simulation process: Decision Stumps, Neural Network, Random Forest and Support Vector Machine. The base classifiers were obtained from the UCI Machine Learning Repository (UCI, 2015) and the London Ambulance Service, NHS Trust (LAS) as illustrated in Table 5. They were selected due to their ability to handle training by reweighting and their applications in managing controlling disease such as breast cancer. In the study, the original version of AdaBoost algorithm i.e. AdaBoost.M1 was applied as the boosting technique (Schapire, 1999). There are two main reasons for using this version of the algorithm:

i. Firstly, to have a generalized experimental set up that is independent of any AdaBoost variant.
ii. Secondly, to have an experimental conclusion that is based on the original version of the algorithm which serves as a primary algorithm for other variants.
iii. Thirdly, the research was based on secondary datasets, errors in datasets and missing data cannot be ruled out.

iv. Fourthly, various parameters were configured to meet the needs of the experimental simulation. Changes to parameter settings can significantly affect the simulation results.

Table 3.1: Experimental Datasets.

Dataset	**Instances**	**Attributes**	**Source**
Breast Cancer	286	10	UCI[2]
Diabetes diagnostic	1000	20	UCI
Credit Card	14980	14	UCI
Hepatitis	155	20	UCI
EEG Eyes Data			
Hypothyroid	3772	30	UCI
Hepatitis			
Labour	57	16	UCI
Supermarket	4625	217	UCI
Thoracic Surgery	470	33	UCI
Primary Tumour			
Unbalanced data	856	33	UCI
Congressional Voting	435	16	UCI
Zoo	101	18	UCI
Incidents 1 data	1200	7	LAS[3]
Incident 2 data	1200	7	LAS

The test involves application of AdaBoost with 10-fold validation and ensembles of 5 to 25 classifiers based on the size of the samples in Table 3.1. The experimental datasets illustrated

[2] UC Irvine Machine Learning Repository

[3] London Ambulance Service

in Table 3.1 compromises of 16 different datasets that were obtained from different sources. The two sources are the UC Irvine Machine Learning Repository (UCI, 2015) and the London Ambulance Service (LAS., 2021). It also illustrated the number of instances and attributes in datasets. Several pre-processing of the datasets were carried out to eliminate anomalies and missing instances to avoid misclassifications.

Four different types of base classifiers were used: decision stump, artificial neural network, random forest and support vector machine. Two different tests were carried out. The first test was based on boosting by reweighting and boosting by resampling. The second test was based on standalone classifiers. In both tests as each classifier is added to the ensemble the following were considered: training accuracy of the ensembles, testing accuracy of the ensembles and time complexity i.e. the time taken to run and finish the execution of the ensembles.

3.4.3 Results and Discussion

Table 3.3 illustrates the statistical performance results and comparison of boosting by resampling and boosting by reweighting. It comprises of 16 datasets in column the dataset columns as illustrated in Table 3.1 and 5 different base classifiers as revealed in section 3.4.2. The base classifiers are namely Decision stump, Neural Network, Random Forest, Support Vector machine and AdaBoost as meta-classifier. In order to meticulously compare the performances of the two methods; different statistical concepts and measures were adopted. These strategies as demonstrated Table 3.1 are is illustrated in Table 3.2 below.

Table 3.2 Different resampling and reweighting strategies

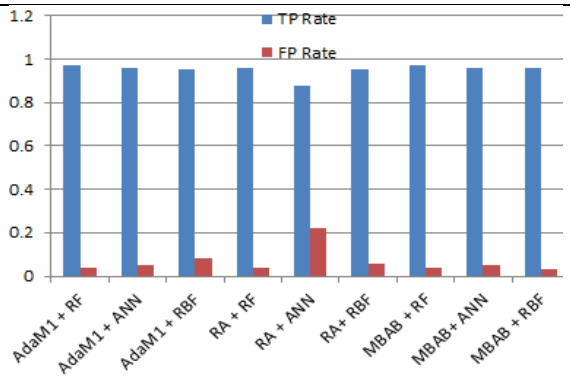
Evaluation metrics	**Description of Evaluation metrics**
C+ sample	Correctly classified with cost-sensitive evaluation by resampling
C- sample	Correctly classified without cost-sensitive evaluation by resampling
C+ weight	Correctly classified with cost-sensitive evaluation by reweighting
C- weight	Correctly classified without cost-sensitive evaluation by reweighing
VR sample	Validation results based on Booting by resampling
VR weight	Validation results based on Booting by reweighting
TE sample	Test results based on Booting by resampling
TE weight	Test results based on Boosting by reweighting
KS sample	Kappa statistics based on Booting by resampling
KS weight	Kappa statistics based on Booting by reweighting
PRC-A weight	Precision Recall curve based on Booting by reweighting
PRC-A sample	Precision Recall curve based on Booting by resampling
ROC-A weight	Receiver operating characteristic – Boosting by reweighting
ROC-A sample	Receiver operating characteristic – based on Boosting by resampling
RMSE weight	RMSE result based on Boosting by reweighing
RMSE sample	RMSE result based on Boosting by resampling

Table 3.3 shows the comparative results of the two methods from the table, it illustrates some of the main performance metrics of the two AdaBoost boosting methods.

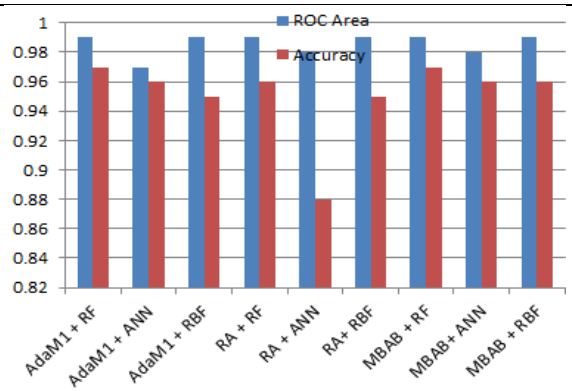
Dataset/Algorithms		Performance comparison of Boosting by Resampling (sample) /Boosting by Reweighting (weight)											
Dataset	Algorithm	C+ sample	C+ weight	C- sample	C- weight	KS sample	KS weight	PRC-A sample	PRC-A weight	ROC-A sample	ROC-A weight	RMSE sample	RMSE weight
		VR | TR	VR | TR	VR | TR	VR | TR								
Breast cancer	DS	72 | 76	70 | 76	28 | 24	30 | 24	0.283	0.256	0.716	0.732	0.682	0.697	0.437	0.430
	NN	71 | 98	71 | 98	29 | 02	29 | 02	.272	.272	.700	.700	.670	.670	.524	.524
	RF	66 | 98	66 | 98	34 | 02	34 | 02	.163	.150	.667	.676	.610	.632	.562	.551
	SVM	73 | 77	73 | 77	27 | 23	27 | 23	.286	.287	.730	.730	.683	.683	.441	.441
	AdaBoost	70 |74	69 |76	30 | 26	31 | 24	.233	.210	.691	.677	.661	.635	.461	.465
Credit data	DS	71 | 77	69 | 76	29 | 23	31 | 24	.210	.169	.754	.740	.737	.732	.426	.431
	NN	71 | 100	71 |99	29 | 0 |1	29 | 1	.318	.318	.693	.693	.672	.672	.525	.525
	RF	74 |100	73|100	26 | 0	27 |0	.335	.303	.767	.768	.750	.732	.482	.508
	SVM	66 |100	66|100	34 | 0	34 |0	-.013	-.013	.592	.592	.517	.517	.578	.578
	AdaBoost	75 |76	72 | 78	25 | 24	28 |22	.359	.268	.783	.750	.762	.730	.426	.442
Diabetes	DS	76 |77	76 |77	24 |23	24 |23	.560	.460	.822	.822	.817	.822	.405	.405
	NN	75 |90	75 | 81	25 | 10	25 | 19	.432	.432	.789	.789	.789`	.789	.432	.432
	RF	73 |100	74 |100	27 |0	26 | 0	.371	.401	.775	.778	.772	.778	.267	.506
	SVM	65 |100	66|100	35 |0	34 | 0	0	-.013	.562	.592	.519	.517	.589	.578
	AdaBoost	75 | 81	75 | 79	25 | 19	25 | 21	.445	.446	.778	.780	.783	.784	.445	.438
EEG Eye Data	DS	67 | 71	67 |70	33 | 29	33 | 30	.310	.308	.704	.705	.716	.717	.463	.463
	NN	55 | 62	55 |56	46 | 38	46 | |44	.072	.072	.558	.558	.554	.554	.500	.500
	RF	95|100	94|100	5 |0	6 |0	.892	.884	.987	.984	.987	.986	.218	.231
	SVM	55 |100	55|100	45|0	54|0	.001	.001	.506	.506	.501	.501	.448	.350
	AdaBoost	69 | 70	70 |70	31 | 30	30 |30	.370	.374	.740	.733	.749	.747	.448	.449
Hepatitis	DS	82 |85	83 |90	18 |15	17 |10	.377	.450	.852	.876	.812	.851	.367	.351
	NN	70 | 100	70 |100	30 |0	30 | 0	.383	.383	.767	.767	.768	.768	.437	.437
	RF	84 | 100	85 |100	6 |0	15 | 0	.464	.531	.880	.884	.831	.850	.390	.371
	SVM	79 | 100	79|100	21 |0	21|0	0	0	.673	.673	.498	.498	.454	.454
	AdaBoost	76 | 97	81 | 97	24 |3	19 | 3	.311	.435	.826	.882	.775	.771	.461	.420
Hypothyroid	DS	93 |93	93 | 93	7 | 7	7 | 7	.032	.032	.966	.965	.989	.990	.121	.122
	NN	96 | 100	96 | 97	4 | 0	4 | 3	.560	.560	.947	.947	.865	.865	.500	.500
	RF	99 |100	99 |100	1 | 0	1 |0	.956	.008	.997	.997	.996	.999	.056	.058
	SVM	93 | 100	93|100	7 | 0	7 | 0	.251	.251	.907	.907	.735	.735	.035	.186
	AdaBoost	95 |95	95 |95	5|5	5 | 5	.588	.595	.973	.974	.987	.987	.118	.117
Labor	DS	89 |100	88|98	11|0	12 | 2	.283	.727	.901	.871	.912	.870	.300	.341
	NN	86 | 100	86|100	14 | 0	14 | 0	.272	.692	.939	.939	.923	.923	.337	.337
	RF	88 | 100	89 |100	12 |0	11|0	.163	.764	.916	.947	.908	.942	.329	.324
	SVM	89 |100	89| 100	11|0	11 |0	.286	.774	.925	.925	.947	.947	.324	.324
	AdaBoost	89 |100	88 |100	11|0	22|0	.233	.727	.872	.871	.882	.870	.328	.341
Primary Tumor	DS	28 |27	29 |29	72 | 73	71 |71	.145	.152	.193	.176	.664	.634	.195	.194
	NN	42 | 90	42 | 91	58 | 10	58 | 9	.338	.338	.423	.423	.779	.779	.194	.194
	RF	43 |86	42 |88	57 | 14	58 | 22	.352	.340	.402	.387	.763	.742	.182	.191
	SVM	39 | 47	39 | 47	61 | 53	61 | 53	.257	.257	.222	.222	.623	.623	.236	.520
	AdaBoost	28 | 27	29 | 29	72 | 73	71| 71	.145	.152	.193	.176	.664	.634	.195	.194
Supermarket	DS	75 | 76	75 |76	25 | 24	25 | 24	.434	.434	.791	.791	.795	.795	.414	.414
	NN	64 | 64	64 | 64	36 |36	36 | 36	.468	.468	.538	.538	.500	.500	.482	.482
	RF	64 | 64	64 | 64	36 | 36	36 | 36	0	0	.538	.538	.500	.500	.481	.481
	SVM	47 | 64	47 | 36	53 | 36	53 | 64	-.001	-.001	.537	.537	.500	.500	.767	.528

	AdaBoost	81 | 79	81 | 80	19| 21	19 | 20	.589	.243	.864	.852	.865	.857	.373	.377
Thoracic surgery	DS	84 | 85	84 |85	16 | 15	16 | 15	0	0	.767	.760	.599	.599	.366	.367
	NN	80 | 98	80 | 95	20 | 2	20 | 5	.062	.062	.771	.771	.554	.554	.431	.431
	RF	43 | 100	42| 100	67 | 0	58 |100	.352	.191	.402	.387	.763	.742	.186	.191
	SVM	81| 92	81 | 93	19 | 8	20 |7	-.023	-.023	.767	.767	.537	.537	.393	.393
	AdaBoost	78 | 85	83 |85	22 | 15	17 | 15	.052	.036	.784	.788	.575	.592	.441	.390
Unbalanced data	DS	98 |100	98| 100	2 | 0	2 |0	0	.026	.977	.974	.971	.773	.129	.128
	NN	98 |100	98 |100	2 | 0	2 | 0	.215	-.006	.978	.973	.607	.526	.128	.132
	RF	98 |100	98 |100	2 | 0	2 | 0	-.006	-.004	.973	.974	.526	.554	.132	.125
	SVM	98 |100	98| 100	2 | 0	2 | 0	-.012	-.012	.972	.972	.490	.490	.024	.155
	AdaBoost	98 | 100	98| 100	2 | 0	2 | 0	.119	-.008	.977	.976	.610	.644	.128	.136
Congressional Voting	DS	98 | 97	98 |98	2 | 3	2 |2	.959	.959	.998	.998	.998	.998	.127	.127
	NN	96 | 100	96| 100	4 | 100	4 |0	.908	.908	.964	.964	.974	.974	.197	.197
	RF	94 |100	95| 100	6 | 0	5 | 0	.880	.903	.983	.984	.983	.983	.231	.206
	SVM	96 | 100	96 |100	4 | 0	4 | 0	.913	.913	.989	.989	.989	.989	.191	.042
	AdaBoost	95 | 100	95 |100	5 | 0	5 | 0	.903	.903	.987	.983	.982	.983	.214	.210
Zoo	DS	56 | 60	56 | 60	44 | 40	44 |40	.378	.408	.552	.532	.849	.830	.338	.271
	NN	96 | 100	96 |100	4 |0	4 |0	.945	.945	.973	.973	.993	.993	.105	.105
	RF	97 |100	86|100	4 | 0	14|100	.948	.814	.981	.923	.998	.988	.098	.214
	SVM	75 | 89	75 | 79	25 | 11	25 | 20	.662	.662	.739	.739	.916	.916	.264	.264
	AdaBoost	59 | 60	61| 60	41 | 40	39 | 40	.432	.448	.533	.534	.834	.830	.309	.294
Incidents 1	DS	78 | 14	79 |14	21 | 86	21 | 86	0	0	.651	.643	.546	.526	.138	.138
	NN	79 |	78 |	21 |	21 |	0	0	.643	.651	.526	.546	.138	.138
	RF	97 | 100	74| 100	4 | 0	26 |0	.948	.027	.981	.659	.998	.587	.098	.163
	SVM	79 | 87	79 | 52	21 | 13	21| 47	0	.008	.509	.646	.637	.527	.148	.143
	AdaBoost	79 | 15	79 | 15	21 | 85	21 | 85	0	0	.651	.643	.546	.526	.138	.138
Incidents 2	DS	15 | 15	15 | 15	85 | 85	85 | 85	.079	.078	.100	.096	.746	.736	.207	.207
	NN	47 |	47 |	53 |	53|	0	0	.315	.315	.579	.579	.184	.184
	RF	47 | 100	45 |100	53 |100	56|100	.005	.0178	.295	.278	.547	.522	.138	.195
	SVM	51 | 33	42 | 49	49 | 67	58 | 51	.472	.023	.554	.280	.855	.522	.203	.215
	AdaBoost	47 | 15	47 | 15	53 | 15	53 | 85	0	0	.596	.315	.319	.579	.185	.067
Average / Performance		73.575	73.375	26.438	26.875	.341	.306	.727	.719	.731	.721	.325	.326
Max /Performance		99	15	85	1	.959	-.043	.998	.096	.998	.333	.767	.042
Min /Performance		15	99	1	85	-.043	.959	0.1	.998	.167	.999	.024	0.594

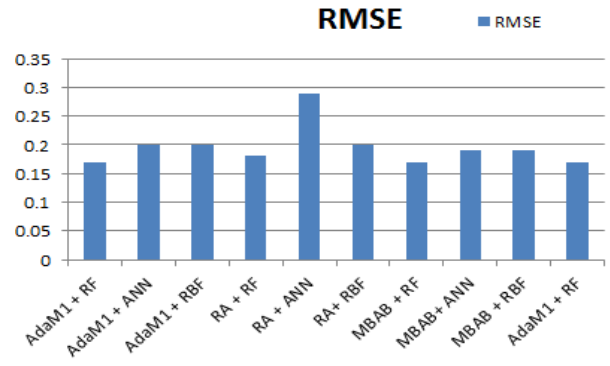
Table 3.3: Empirical Results for Boosting by Reweighting and Boosting by Resampling for all Model Types.



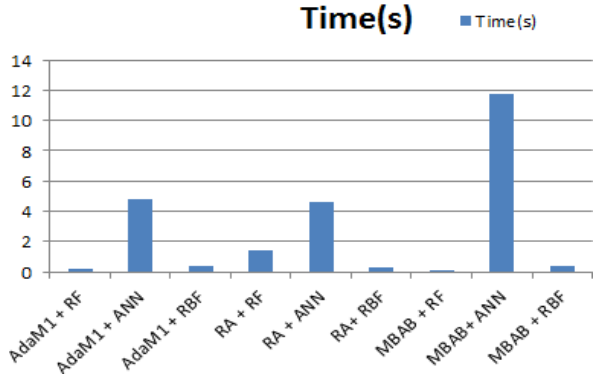
Vertical axis: TP Rate/FP Rate



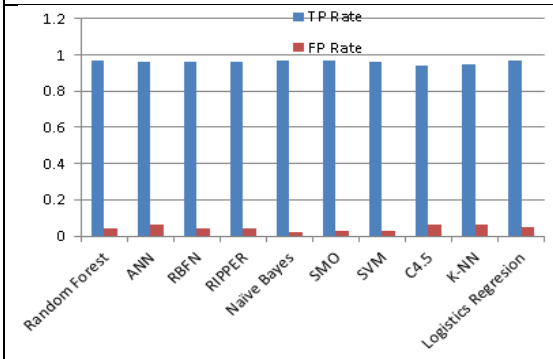
Vertical axis: ROC Area/Accuracy



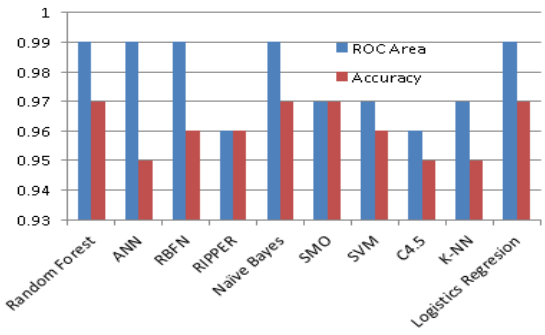
Vertical axis: RMSE



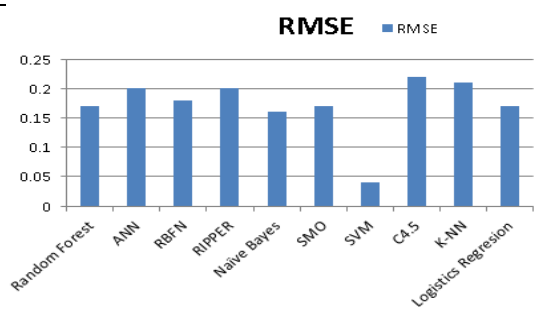
Vertical axis: Time(s)



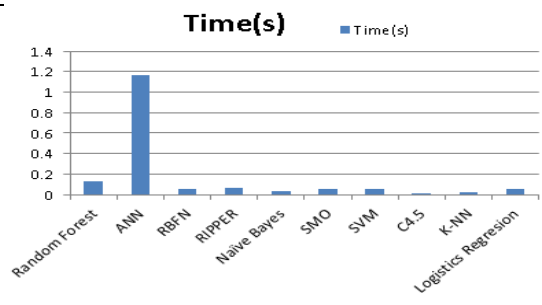
Vertical axis: TP Rate/FP Rate



Vertical axis: ROC Area/Accuracy



Vertical axis: RMSE



Vertical axis: ROC Area/Accuracy

Figure 3-4: Classification Comparisons - Standalone Algorithms and Ensemble Classifiers.

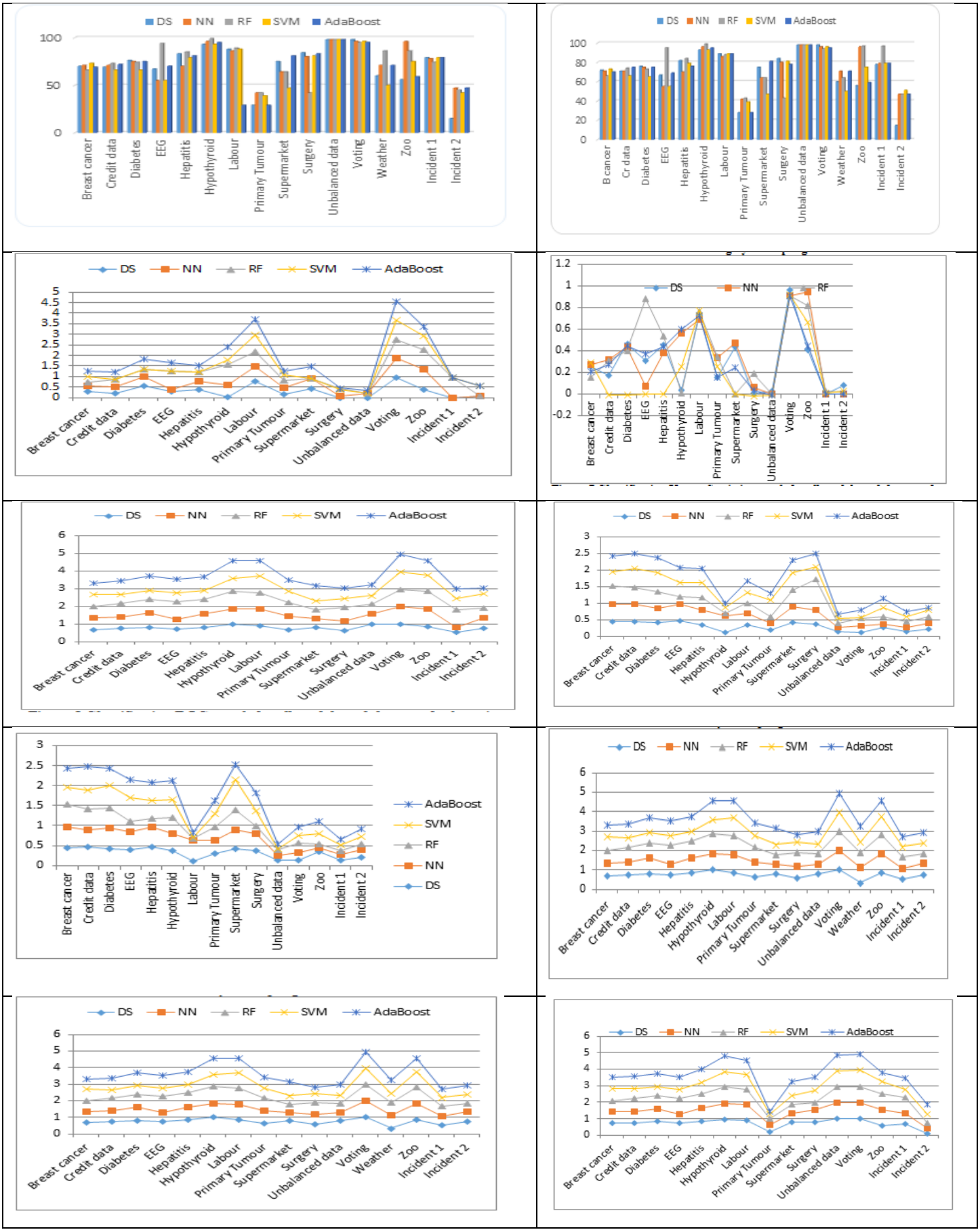


Figure 3-5: Empirical Comparison of Boosting Implementation Methods.

Table 3.4: Summary Comparison of Reweighting and Resampling Methods.

Boosting methods		
Measure	**Reweighting**	**Resampling**
Accuracy	73	74
KS	0.316	0.341
PRC	0.719	0.727
ROC	0.721	0.723
RMSE	0.326	0.325

Table *3.4* compares the summary performance for boosting reweighting and boosting resampling based on accuracy, KS, PRC, ROC and RMSE. As can be seen from Table *3.4*, in summary these results show that in both methods the average of the correctly classified and incorrectly classified are comparatively the same. As illustrated in the table the performance accuracy are 73% and 74% for boosting reweighing and boosting resampling respectively. However, the average values of the RMSE in both methods are slightly different, these are 0.326 and 0.325 for boosting reweighing and boosting resampling respectively. Similarly, as can be seen from the table the KS, PRC and ROC values are also slightly different.

The results in Table *3.4* further shows that the performance of boosting by reweighting and boosting by resampling are independent of the datasets used during the investigation. In comparison, the training accuracy and testing accuracy of the ensembles are significantly the same. Therefore, there is no correlation to the performance accuracy.

More importantly, as can be seen from the comparison data in Table 3.3 and graphical illustration in Figure 3-5, the experimental investigation shows that the complexity of the chosen ensemble technique, boosting method and base classifier do not necessarily lead to a better performance of the models.

3.5 Experimental Investigation: Standalone and Ensemble Models

The emphasis of the investigation in this section is to understand the performance accuracy of standalone and ensemble techniques. To understand this, accuracy of the standalone algorithms and time complexity i.e. the time taken to run and finish the execution of the standalone algorithms were also measured.

Study shows that in many cases standalone models have been applied to estimate the prediction outcome of many systems, however, many have resulted in a lower prediction accuracy than expected. Therefore, in addressing this issue, a number of studies have focused on ensemble methods (Kazienko, et al., 2013; Zhukova, et al., 2017; Hamze-Ziabaria & Bakhshpoorib, 2017; Vašát, et al., 2017; Yousefi, et al., 2018) in order to reduce the variance (Polikar, 2012) and improve prediction accuracy the models.

To investigate and compare performance accuracy of ensemble techniques and standalone models: two experimental investigations were developed bas C4.5, K-NN and Logistics Regression on Breast Cancer Survivability samples.

The purpose of using breast cancer survivability samples is because of the death rates and negative societal impact of the disease globally most especially among women of all ages (Kwon & Lee, 2016; Society, 2015; Fitzmaurice, et al., 2015; Youlden, et al., 2014; WHO, 2017; WHOa, 2021; WHOb, 2020). Therefore, it necessitates the need to develop a non-invasive model to predict the survivability of the disease as part of this thesis to address the problem.

For the ensemble model performance investigation, Random Forest, Radial Basis Function Network and Artificial Neural Network algorithms were applied as base learners. The

following ensemble models: AdAM1, RA and MBAB were applied as meta-classifiers to train the base learners.

Likewise, for the performance investigation of the standalone models ten different standalone classifiers were constructed on the same breast cancer datasets. The standalone algorithms applied in building the models are: Random Forest ANN, RBFN, RIPPER, Naïve Bayes, SMO, SVM, C4.5, K-NN and Logistics Regression.

3.5.1 Results and Discussion

The performance results for the ensemble and the standalone classifiers are as illustrated in Table 3.5 and Table 3.6 respectively and graphically illustrated in Figure 3-4.

Table 3.5 Performance Comparison among the Committee of Classifiers

Metrics / Models	Accuracy	RMSE	TP Rate	FP Rate	Precision	Recall Curve	ROC Area	PRC Area	Time(s)
AdaM1 + RF	0.97	0.17	0.97	0.04	0.97	0.97	0.99	0.99	0.20
AdaM1 + ANN	0.96	0.2	0.96	0.05	0.96	0.96	0.97	0.97	4.77
AdaM1 + RBFN	0.95	0.2	0.95	0.08	0.96	0.95	0.99	0.98	0.41
RA + RF	0.96	0.18	0.96	0.04	0.97	0.96	0.99	0.99	1.44
RA + ANN	0.88	0.29	0.88	0.22	0.9	0.88	0.98	0.98	4.61
RA+ RBFN	0.95	0.2	0.95	0.06	0.95	0.95	0.99	0.99	0.28
MBAB + RF	0.97	0.17	0.97	0.04	0.97	0.97	0.99	0.98	0.14
MBAB+ ANN	0.96	0.19	0.96	0.05	0.96	0.96	0.98	0.98	11.73
MBAB + RBFN	0.96	0.19	0.96	0.03	0.96	0.96	0.99	0.98	0.37

Table 3.6 Performance Comparison among Standalone of Classifiers.

Metrics / Algorithms	Accuracy	RMSE	TP Rate	FP Rate	Precision	Recall Curve	ROC Area	PRC Area	Time(s)
Random Forest	0.97	0.17	0.97	0.04	0.97	0.97	0.99	0.99	0.13
ANN	0.95	0.2	0.96	0.06	0.97	0.96	0.99	0.98	1.17
RBFN	0.96	0.18	0.96	0.04	0.96	0.96	0.99	0.98	0.05
RIPPER	0.96	0.2	0.96	0.04	0.96	0.96	0.96	0.95	0.06
Naïve Bayes	0.97	0.16	0.97	0.02	0.97	0.97	0.99	0.99	0.03
SMO	0.97	0.17	0.97	0.03	0.97	0.97	0.97	0.96	0.05
SVM	0.96	0.04	0.96	0.03	0.96	0.96	0.97	0.95	0.05
C4.5	0.95	0.22	0.94	0.06	0.94	0.94	0.96	0.94	0.01
K-NN	0.95	0.21	0.95	0.06	0.95	0.95	0.97	0.96	0.02
Logistics Regression	0.97	0.17	0.97	0.05	0.97	0.97	0.99	0.99	0.05

In the case of the ensemble models, referring to Table 3.5 and Figure 3-4, we can see that the AdaM1 + RF, RA + RBFN, MBAB+RF and MBAB + RBFN models produced prediction accuracy of 97%, 95%, 95%, 97% and 96% respectively with execution time of 0.20s, 0.28s, 014s and 0.37s, respectively. However, ANN ensembles: AdaM1 + ANN, RA + ANN and MBAB+ ANN models produced prediction accuracy of 96%, 88% and 96% respectively with execution time of 4.77s, 4.61s and 11.73s respectively.

The results further shows that the complexity and topology of a chosen base classifiers does not automatically leads to improved ensemble predictive results of the models. Contrarily, in some cases it produces the worst performance results despite high execution time and CPU resources required for training. As can be seen in Table 3.5, RA + ANN model produces worst prediction accuracy of 88% and execution time of 4.61s.

The performance results of the standalone models are as details in Table 3.6 and graphically illustrated in Figure 3-4. As can be seen in Table 3.6 random forest, logistic regression, naïve Bayes and SMO all have the same performance accuracy of 97%, with different execution times. However, it takes Random Forest and Naïve Bayes models 0.13s and 0.03s to execute the sample while it takes SMO and Logistic regression 0.05s respectively to execute the same sample. It can be seen from the data in Table 3.6 that it takes ANN model 1.17s with a prediction accuracy of 95%. As illustrated in Figure 3-4 it further shows that complexity of the standalone ANN model with execution time of 1.17s and predictive accuracy of 95%. The model did not outperform any of the simple standalone models as can be seen in Table 3.6

3.6 Chapter Summary

In this chapter two distinct empirical investigations to establish the performance of the various standalone and ensemble techniques have been carried out. These are:

I. Pragmatic comparison of boosting methods; boosting resampling and boosting by reweighting implementation techniques.
II. Experimental performance and comparison of standalone and ensemble methods.

The results from both experimental investigations provide answers to the Hypothesis 1 and Hypothesis 2 of this thesis. The results presented in this chapter shows that boosting by sampling method performs relatedly better than boosting by reweighting method. The complexity and topology of algorithm models (in terms of time and space) does not necessarily make them perform better than the simple models.

It the next chapter the proposal, development, discussion and testing of ensemble EKF-RBFN-ADA concept is presented. A model that integrates there concepts namely: EKF, RBFN, and AdaBoost as a framework to enhance the predictive performance of ensemble models.

Chapter 4: EKF-RBFN-ADA Ensemble Models

4.1 Introduction

In the previous chapter various experimental investigations to compare the performance of boosting methods were carried out. Performances of standalone and ensemble algorithms were also considered and compared which provide answer to Hypothesis 1 and Hypothesis 2 of the thesis that were derived from the literature reviews.

A number of reports show that hybrid modelling concepts are widely used to predict changes of process variables and to enhance the overall performance of such systems (Zu, et al., 2012; Olsson, et al., 2001; Chou, et al., 2013; Aly, 2020; Javid, et al., 2020; Zain & Tokhi, 2004). The procedures improve the performance of machine learning methods by leveraging and combining the strengths and weaknesses of several models to obtain a better and stronger classifier than averaging the models.

The focus of this phase is to address Objective 2, the integration of EKF, RBFN and AdaBoost algorithms to combine the weak classifiers that could provide relatively parsimonious framework compared to existing ensemble methods.

However, review of the literature shows that no previous research especially on how to develop hybrid ensemble models that amalgamate the three methods that is EKF, RBFN and AdaBoost models, has been considered as an alternative solution to improve predictive performance of ensemble models.

In this chapter a brief outline of the proposed EKF-RBFN-AdaBoost models that is based on consolidation of Extended Kalman Filter (EKF), Radial Basis Function Networks (RBFN) and AdaBoost modes, is discussed. Doing this enables us to address the current inefficient

performance of ensemble models. In the proposed model, EKF is applied as training algorithm to train and optimize the parameters of Radial Basis Function Networks (RBFN) to produce EKF-RBFN object as output. Then AdaBoost as a meta-algorithm accepts the EKF-RBFN outputs as input prototypes which are further trained as committee of weak classifiers by AdaBoost to output a stronger classifier. The methodology extends the application of hybrid modelling concept to non-linear dynamical systems that is approximated using the first or second order derivatives of the EKF and RBFN models.

The following sections and subsections of this chapter focus on how to develop and integrate the three models as an ensemble model in order to improve its prediction accuracy with focus on breast cancer survivability and diabetic models. Firstly, a brief background and related work on Cancer and Diabetic models are presented. Secondly, the theory of ensemble modelling is described, and thirdly the proposed model is discussed followed by the model's performance analysis and conclusion is outlined.

4.2 Related Cancer and Diabetic Models

Review shows that ensemble algorithms are essentially iterative, notwithstanding their results are inconsistent and not as accurate as it should be. For instance, the application of ensemble algorithms in early prediction of breast cancer and diabetes diagnosis which are the two common diseases that affects a lot of people requires algorithms with high predictive accuracy and reliability (Thomas, 2019; Kelly, et al., 2019; Ghassemi, et al., 2020).

Breast cancer has been classified as one of the most common causes of cancer related death mostly among women in the world in the past years. In the USA alone, in 2015 an estimated 231,840 new cases of invasive breast cancer were diagnosed among women and 60,290 additional cases of in-situ breast cancer (American-Cancer-Society, 2015; Adegoke, et al.,

2017). Similarly, in the UK over 55,222 women were diagnosed with new cases of the disease in 2014 which amounted to 11, 433 deaths (UK, 2018) and the ailment reached 25.2% of women worldwide (Kwon & Lee, 2016).

There have been several investigations that show that the disease is also a looming epidemic in the developing countries where advanced techniques for early detection and treatments are not readily available (Formenti, et al., 2012).

Similarly, "Diabetes is a chronic progressive disease that is characterized by elevated levels of blood glucose. It has been argued that diabetes of all types can lead to complications in many parts of the body. Therefore, it can increase the overall risk of dying prematurely" (WHO, 2016). According to the British Heart Foundation "the increasing number of people suffering from the epidemic could trigger a 29% rise in the number of heart attacks and strokes linked to the condition by 2035" (BHF, 2018; ITV, 2018). Recent analysis show that about four million people in the UK have diabetes with condition accounting for 10% of all NHS spending (BBC, 2018).

In recent years researchers have investigated a variety of approaches in data mining using different ensemble techniques in predicting probable events based on historical datasets. One of the key challenges is the choice of the base classifiers and the appropriate loss functions that go with it. The goal of any ensemble algorithm is to minimize error rate in order to achieve accuracy and improve reliability. Despite the successful research efforts and application of ensemble methods recent work shows that the problem with prediction accuracy, speed and computational cost are still puzzling tasks. Therefore, the development of reliable ensemble models that can be applied for efficient medical diagnosis, incidents management and execution of automated technologies that are decision based, and in some

cases life dependent are highly essential. To address the issue of prediction accuracy, reliability and to extend the applications of ensemble algorithms, the research proposed a new model that bridges the performance potentials of RBFN, EKF and AdaBoost algorithms.

4.2.1 Breast Cancer Survivability Models

Medically, breast cancer can be detected early during screening examinations through mammography or after a woman notices an unusual lump (American-Cancer-Society, 2015) in her breast. Owing to advancement in technology and availability of patient medical records, computer aided diagnosis cancer detection systems have been developed to detect and thus control the spread of the disease. However, such systems rely on pattern recognition algorithms that are used to process and analyse medical information of images obtained from mammograms for diagnostic and decision making (Weedon-Fekjær, et al., 2014; Sapate & Talbar, 2016).

Different algorithms have also been proposed to extract relevant patterns from patients' breast cancer ailments; for instance Yang et al (Yang, et al., 2013) came up with a genetic algorithm that identify the relationship between genotypes that can lead to cancer cases using mathematical analysis. Similarly, McGinley et al (McGinley, et al., 2010) applied Spiking Neural Networks algorithm as a novel tumour classification method in classifying cancer tumours as either benign or malignant.

Recently, a deep learning-based approach has been applied to high dimensional, high-volume, and high-sparsity medical data to identify critical casual attributions that might affect the survival of a breast cancer patient. More recently, a deep learning-based approach has been applied to high dimensional, high-volume, and high-sparsity medical data to identify

critical casual attributions that might affect the survival of breast cancer patients (Chen, et al., 2021).

In another approach (Pak, et al., 2015) proposed a breast cancer detection and classification in digital mammography based on Non-Subsampled Contourlet Transform (NSCT). The authors proposed a Super Resolution to improve the quality of digital mammography images. The authors then applied AdaBoost algorithm to determine the probability of a disease being a benign or malign cancer. Likewise, in breast mass cancer classification (Xie, et al., 2015) the authors used computer-aided diagnosis (CAD) system for the processing and diagnosis of breast cancer.

4.2.2 Diabetes Diagnostic Models

In their study Alghamdi *et. al.* using SMOTE and ensemble techniques carried out experimental work applying a number of algorithms to establish and compare their performances in predicting diabetes-based data obtained from patients' medical history (Alghamdi, et al., 2017). The model comprises of ensemble-based predictive method that uses 13 out of the 62 available classified attributes. The selected attribute for the model depends on clinical importance, multiple linear regression (MLR) and the Information Gain (IG). The authors reported an accuracy performance of 89% for G1/G2 attributes and accuracy (AUC) of 0.922 for the ensemble method. Similarly, in (Zheng, et al., 2017), the authors proposed a framework that identifies type 2 diabetes using patient's medical data. They utilized various classification models that extract features to predict identification of T2DM in datasets. According to the authors, the average results of the framework was 0.98 (UAC) compare with other algorithms at 0.71. To validate whether there is a connection between diabetes mellitus and glaucoma chronic diseases, in their work (Apreutesei, et al., 2018) applied a simulation

technique constructed using artificial neural networks on clinical observations datasets. According to the authors the model was able to predict an accuracy of 95%.

In another study (Barakat, et al., 2010) the authors proposed a multi-purpose model for the diagnosis and prediction of diabetes using support vector machines algorithm. According to the authors, the results of the model show a prediction accuracy of 94%.

4.3 Ensemble Modelling

As previously highlighted, the purpose of ensemble modelling is to improve prediction accuracy by producing a stronger learner through the combination of multiple weak models. Therefore, the overall performance of the combined ensemble weak classifiers is boosted by taking the advantage of the dependency and diversity among the base learners. The concept penalize instances that are not correctly classified with higher weights. Therefore, it awards instances that are correctly classified with little or no weights as a performance compensation. In general, ensemble model consists of several related ensemble members that are combined with their ensemble weights into a predictive model. The purpose of this approach instead of the conventional or standalone classifier method that is based on a single pool of dataset is to decrease variance and bias, and therefore improve predictions accuracy. In this session, more emphasis is laid on the concept and formulation of ensemble models.

4.3.1 Ensemble Model Formulation

An ensemble model with m weak classifiers can be described as shown in Equation (4.1):

$$Y = w_1 y_1 + w_2 y_2 + w_3 y_3 \ldots\ldots + w_m y_m \quad (4.1)$$

where Y is the output of the ensemble models, $y_1, y_2, y_3 \ldots, y_m$ are the prediction output of ensemble members and $w_1, w_2, w_3, \ldots, w_m$ are the corresponding ensemble weights of ensemble members $1, 2, 3, \ldots . . and\ m$, respectively. The accuracy of equation (4.1) is determined by the ensemble weights of the weak classifiers which are determined during training. In ensemble modelling the tasks is to minimize the difference between the predicted outputs and the actual class label, the right hand and left hand sides of Equation (4.1) respectively.

Therefore, we can re-arrange Equation (4.1) in the form of an objective function that describes the classification error such that:

$$E = \sum_{i=0}^{m} (y_i - \bar{y})^2 \tag{4.2}$$

where E is the classification error, y_i is the prediction i instance and $\bar{y}$ is the mean of all the target in the group. As mentioned, in any modelling the goal is to improve the prediction accuracy by reducing the error, E. Therefore, if we introduce weights and biases into Equation (4.2), it can be rearranged and expressed the equation in the form of a cost function as in Equation (4.3) such that:

$$F(w, b) = \sum_{i=0}^{m} (y_i - \bar{y})^2 \tag{4.3}$$

Thus in Equation (4.3), the cost function shows how accurate the prediction error is, if it is high it shows poor prediction performance. While on the other hand, a low value indicates a higher performance accuracy. The predicted output in Equation (4.3) can consequently be expressed as shown in Equation 4.4, such that:

$$\bar{y} = \sum_{j}^{m} W * X + b \tag{4.4}$$

Equation 4.4 can also be rearranged such that the error can be expressed in form of Δ (delta) that is change in the input and output values in Equation (4.4) as expressed in Equation (4.5), such that:

$$\Delta\bar{y} = \sum_{j}^{m} \Delta W * \Delta X + \Delta b \tag{4.5}$$

In Equation 4.4 and Equation 4.5 ensemble weights ($W\ and\ \Delta W$) and the biases (b and Δb) play important role on the classification and performance accuracy of ensemble models.

The weights and biases are the learnable parameters of machine learning models. For instance in the neural networks the weights control the signal between two neurons. It determines the influence of the input on the output while the biases are additional input to the next layer (Géron, 2018; Schmidt, et al., 2019) and essential in optimizing ensemble model analysis and performance (Shahhosseini, et al., 2019; Kotu & Deshpande, 2019).

Assuming that two ensemble members that are unbiased with outputs of y_1 and y_2, and independent variances σ_1 and σ_2 such that:

$$\beta(Y - y_1) = \beta(Y - y_2) = 0$$

$$COVAR(y_1, y_2) = 0$$

Therefore, an ensemble model with two members can be shown as

$$Y' = w_1\, y_1 + w_2\, y_2$$

where w_1 and w_2 are the ensemble weights. For w_1 and w_1 that are non-negative and normalised ensemble weights, such that $w_1 + w_2 = 1$. Therefore, the bias of the model can be expressed as:

$$\beta(Y - Y') = \beta((w_1 + w_2)Y - w_1\, y_1 - w_2\, y_2)) = 0$$

This can further be rearranged as

$$\beta(Y - Y') = w_1\, \beta(Y - y_1) + w_2\, \beta(Y - y_2) = 0$$

Hence, the variance, Y' can be expressed as:

$$VAR(Y') = VAR(w_1\, y_1 + w_2\, y_2) = w_1^2\, \sigma_1 + w_2^2\, \sigma_2 = w_1^2\, \sigma_1 + (1 - w_1)^2 \sigma_2$$

Taking the derivate of $VAR(Y')$ with respect to w_1 and w_2 and setting the result to zero respectively will minimize $VAR(Y')$ such that:

$$\frac{\partial VAR(Y')}{\partial w_1} = \frac{\partial VAR(Y')}{\partial w_2} = 0$$

Therefore,

$$w_1 = \frac{\sigma_2}{\sigma_1 + \sigma_2} \text{ and } w_2 = \frac{\sigma_1}{\sigma_1 + \sigma_2}$$

Alternatively,

$$w_2 = 1 - w_1 = \frac{\sigma_1}{\sigma_1 + \sigma_2}$$

Therefore, whenever $\sigma_2 > \sigma_1$ then:

$$VAR(Y') - \sigma_1 = w_1^2\, \sigma_1 + (1 - w_1)^2 \sigma_2 - \sigma_1 \quad (4.6)$$

This shows that when the output of Equation (4.6) is negative, then the performance of the ensemble model is better than performance of individual members of the classifier. However,

if the output value of the model is positive, then other member of the classifiers should be used as estimator instead of $Y\ or\ Y'$.

The above variance-bias analysis shows that ensemble weights play crucial impact on the performance of machine learning concepts that are based on ensemble modelling (Schapire & Freund, 2014; Cover & Thomas, 2006). It is therefore significant to effectively optimize ensemble weights in order to enhance their predictive performance efficiency, robustness and reliability.

4.4 Hybrid EKF-RBFN-AdaBoost Models

In this section and other subsections, a brief outline theory and the structures of the three models: EKF, RBFN and AdaBoost that were integrated to form the framework of the proposed hybrid models are discussed. The development of the hybrid model that was centred on differentiable based-process models are also briefly discoursed.

4.4.1 RBFN Models

RBF network implements an input–output mapping using a linear combination of radially symmetric functions (Moody & Darken, 1989). The algorithm have characteristics similar to those of back propagation networks. The algorithm can be viewed as an alternative to the multi-layer perceptron neural network (MLPNN) for non-linear modelling (Nabney, 2002).

They do a nonlinear, through localized, transformation and a weighted linear combination. It has three layers, with feedforward connections between the nodes, as illustrated in Figure *4-1* (Galar & Kumar, 2017).

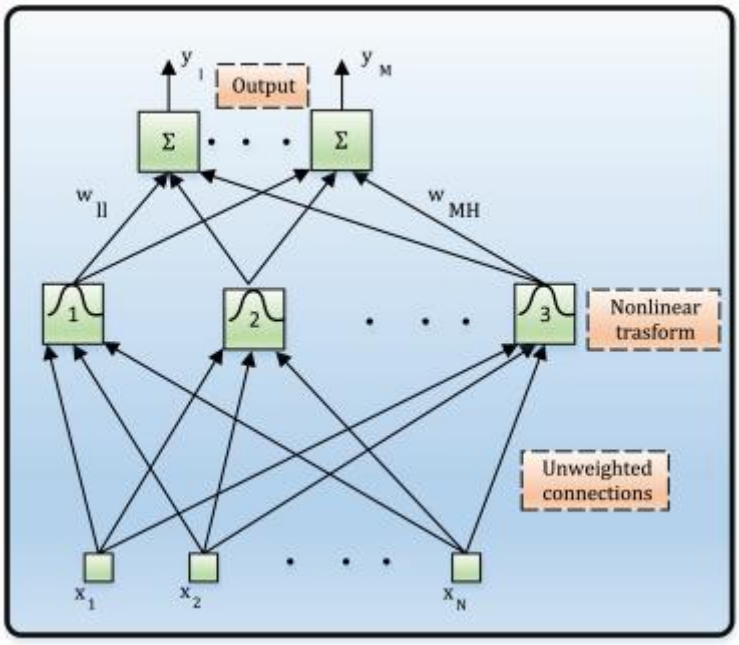


Figure 4-1: Radial Basis Function Network *(Galar & Kumar, 2017)*.

Several algorithms have been proposed that are used by researchers to train RBF Networks. Rather than picking random data points, a principled clustering technique to find a set of RBF centres which accurately reflect the distribution of the data points has been used. This technique include unsupervised methods such as the K-means clustering (Qiao, et al., 2016; Xing, et al., 2015) and self-organizing maps (Yamashita, et al., 2010; Chang, et al., 2007).

Some researchers have also used supervised methods such as Particle Swarm Optimization (PSO) (Kelwade & Salankar, 2016; Wang, et al., 2015) and Gradient Descent (Malathi & Suresh, 2014; Soni, et al., 2015), at the same time others have used Artificial Bee Colony (Kurbban & Besdok, 2009) algorithms to determine the parameters of the network. However, in this research EKF is used to train RBF networks in order to optimize the network parameters. This is because EKF is regarded as an accurate function estimator that converges faster during training (Rhudy, 2015; Rivals & Personnaz, 1998). The algorithm can also handle small training samples and missing data effectively compare to other available methods. As previously

described after the training of the network with EKF, AdaBoost is employed as a meta-technique to train the EKF-RFFN model as an ensemble of RBF networks.

4.4.2 Kalman Filter Models

Kalman Filter (Kalman, 1960; Li, et al., 2015) is an optimal estimator algorithm that deduces unknown values of interest from inaccurate and uncertain or inaccurate observations. Even though the filter was originally developed as a recursive solution to the discrete data linear filtering problem, it has been used to estimate linear system models with additive independent white noises. The filter uses a number of measurements observed over time that contains noises and other inaccuracies which it filters to predict the future behaviour of a system based on the system's past behaviour, taking into consideration the environmental constraints of the system. Therefore, the filter minimizes the mean square error of the estimate.

The extended Kalman filter (EKF) is the nonlinear version of the Kalman Filter which linearizes the estimate of the current mean and covariance Lacey shows that the filter has been considered as a standard in the theory of nonlinear state estimation, navigation systems, data prediction tasks (Lacey, 2020) and other related problems by several researchers over the last decades.

As demonstrated in the previous section the process of calculating the ensemble weights can be considered as a discrete and sequential estimation problem (Kalman, 1960; Lacey, 2020).

Therefore, the state-space model can be derived in terms of the weights and outputs of the ensemble model described in Equation 4.1 can be sequentially represented as follows in Equation (4.7) and Equation (4.8), respectively.

$$w_1 = w_{k-1} + \omega_{k-1} = \begin{bmatrix} w_{k-1,1} \\ w_{k-1,2} \\ \vdots \\ w_{k-1,n} \end{bmatrix} + \omega_{k-1} \tag{4.7}$$

$$Y_k = Y_k' + v_k = [y_{k,1}\, y_{k,2} \cdots\cdots y_{k,n}] + \begin{bmatrix} w_{k,1} \\ w_{k,2} \\ \vdots \\ w_{k,n} \end{bmatrix} + v_k \tag{4.8}$$

Where

w_k is the ensemble weights at time k,

Y_k' is the output of the model at time k,

Y_k is the measurent at time k,

ω_{k-1} is process noise, and

v_k is the measurement noise

Assuming that the process noise ω_{k-1}, and the measurement noise v_k, are Gaussian white noises with covariance Q_{k-1} and R_k, respectively, therefore, KF can be applied to estimate and optimize the training weights in Equation (4.1) and Equation (4.2). In addition, KF can be recursively applied as a sequential ensemble method that represents the concept of sequential ensemble modelling described in this session as illustrated in Equation (4.7) and Equation (4.8). The recursive nature of Kalman Filter algorithm as sequential ensemble technique is as illustrated in Figure 4-2.

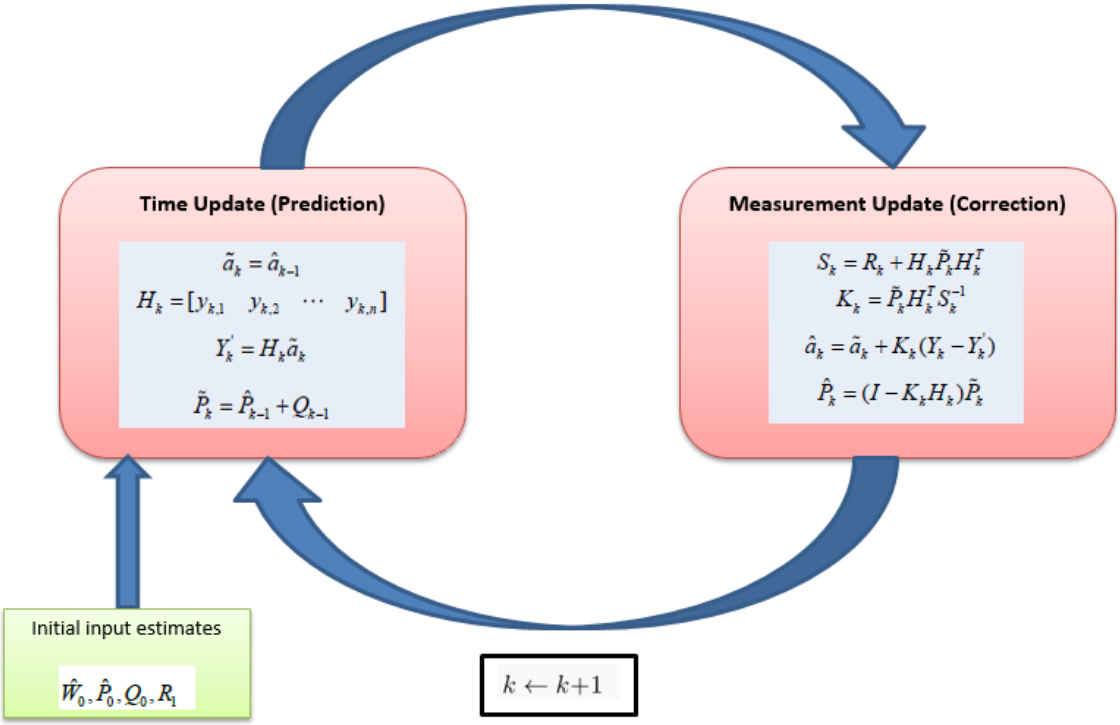


Figure 4-2: Kalman Filter Recursive Algorithm – A Sequential Ensemble Technique

The bias and the variance of Y_k' in Equation 4.8 can be calculated as shown below assuming that $E(a_{k-1} - \hat{a}_{k-1}) = B_{k-1}$

Bias of Y_k':

$$E(Y_k - Y_k') = E(H_k a_k + v_k - H_k \hat{a}_k)$$

$$= H_k . E(a_k - \hat{a}_k)$$

$$= H_k . E(a_{k-1} + \mu_{k-2} - \hat{a}_{k-1})$$

$$= H_k . E(a_{k-1} - \hat{a}_{k-1})$$

$$= H_k . B_{k-1}$$

Variance of Y_k':

$$VAR(Y_k - Y_k') = E[(Y_k - Y_k')(Y_k - Y_k')^T]$$

$$= E\{[H_k(a_k - \hat{a}_k) + v_k][H_k(a_k - \hat{a}_k) + v_k]^T\}$$

$$= H_k\, E[(a_k - \hat{a}_k)(a_k - \hat{a}_k)^T] H_k^T + R_k$$

$$= H_k \hat{P}_k\, H_k^T + R_k$$

where,

$\hat{P}_k = \hat{P}_{k+1} + Q_{k-1}$, and $\hat{P}_{k+1}$ is the trace which is minimised at time, $t = k - 1$. This makes the trace of Y_k' to be minimum and therefore can be visualised as a form of sequential Mean Square estimation.

4.4.3 Optimization of RBFN using Kalman Filter

As demonstrated in the previous section the optimization of the ensemble weights is a type of discrete data filtering problem. Therefore, it is possible to use Kalman Filter to optimize the weight matrix problems. Also, the training error of the ensemble model can be treated as a least square minimization problem. The derivation of extended Kalman Filter (EKF) as a sequential ensemble method are widely available in the literature (Wan & Van Der Merwe, 2000; Haykin, 1996; Ribeiro, 2004).

Therefore, in this session importance is laid on how Kalman filter can be applied to optimize the training parameters of Radial Basis Function Networks in order to improve the algorithm's prediction performance.

Assuming a non-linear finite dimension discrete time system (Chernodub, 2014; Haykin, 2008; Simon, 2013; Simon, 2002) we can represent the state and measurements as

$$\theta_{k+1} = f(\theta_k) + \omega_k \tag{4.9}$$

$$y_k = h(\theta_k) + v_k \tag{4.10}$$

where the vector θ_k is the state of the system at time k, ω_k is the process noise, y_k is the observation vector, v_k is the observation noise and $f(\theta_k)$ and $h(\theta_k)$ are the non-linear vector

functions of the state and process respectively. If the dynamic state $f(\theta_k)$ and process $h(\theta_k)$ in equations 4.9 and 4.10 are assumed known, then EKF can be used as the standard method of choice to approximate maximum likelihood estimation of the state θ_k (Wan & Van Der Merwe, 2000). Therefore, the optimization of RBFN with weight W and the prototype v_j as a weighed least-square minimization (WLSM) problem can be viewed.

The error vector can also be viewed as the difference between the RBFN outputs and the expected target values. Therefore, the optimization problem of RBFN can be represented using Kalman filter algorithm by letting the output of the weight W and the elements of the prototype v_j represent the state of a nonlinear system and the output of the RBFN network respectively (Chernodub, 2014; Simon, 2002). Therefore, the problem addressed by EKF is to find an estimated values for $\hat{\theta}_{n+1}$ of θ_{k+1} given that $y_j(j = 0, \dots, k)$. Hence, the estimated value $\hat{\theta}_n$ can be obtained using recursion expressions as

$$\hat{\theta}_k = f\left(\hat{\theta}_{k-1} + K_k\left[y_k - h\left(\hat{\theta}_{k-1}\right)\right]\right) \quad (4.11)$$

$$K_k = P_k K_k (R + H_k^T P_k H_k)^{-1} \quad (4.12)$$

$$P_{k+1} = F_k(P_k - K_k H_k^T P_k)F_k^T + Q \quad (4.13)$$

$where\ K_k$ is the Kaman Gain, P_k is the covariance matrix of the estimation error, θ_{k+1} is state estimation, Q is the process noise covariance matrix for ω_k and R is the measurement noise covariance matrix.

4.4.4 Analytical Representation of RBFN

Assuming a RBFN model with m inputs, c prototypes, n outputs, using y to denote the target vector for the network output, $\hat{y}$ (otherwise $h(\hat{\theta}_k)$) to represent the actual output of the

model at the kth iteration of the optimization model. The optimization of the weight w and the prototypes v_j can be viewed as a weighted least-squares minimization task.

The error vector is the difference between the RBF Network outputs and the expected output values, such that:

$$y = [y_{11} \cdots y_{1M} \cdots y_{n1} \cdots y_{nM}]^T$$

$$\hat{y} = [\hat{y}_{11} \cdots \hat{y}_{1M} \cdots \hat{y}_{n1} \cdots \hat{y}_{nM}]^T$$

where y and $\hat{y}$ are vectors that consists of nM elements: n is the dimension of the RBFN and M is the number of training samples. To use KF to represent and optimize RBFN parameters firstly, let the elements of the weight matrix W and the elements of the prototypes y_i constitute the state of a nonlinear system. Secondly, let the output of the RBF network constitute the input of the nonlinear system to which the KF is applied. Therefore, the state of the nonlinear system of the RBFN can therefore be characterised as:

$$\theta = [w_1^T \cdots w_n^T \quad v_1^T \cdots v_c^T]^T$$

It is important to note that the vector θ is a linear array of all $(n(c+1)+mc)$ of the RBFN parameters. Therefore, the non-linear model to which Kalman can be applied is as expressed in Equation (4.5) and Equation (4.6) as

$$\theta_{k+1} = \theta_k + \omega_k \tag{4.14}$$

$$y_k = h(\theta_k) + v_k \tag{4.15}$$

where $h(\theta_k)$ is RBFN's nonlinear mapping between its parameters and its outputs, ω_k and v_k are the process and measurement noises (Simon, 2013) respectively.

4.5 Proposed EKF-RBFN-Ada Model Description

In this subsection and the next, a description of the various components of the proposed model is provided. One of the key benefits of the model is that parameters in the ensemble members are sequentially adjusted during training as iterated above. Therefore, using Kalman filter as a sequential algorithm does not necessarily depends on large data or missing data to make adequate prediction. However, it only needs the past state and current state of the system to make a prediction of the next state of the system. This is unlike the conventional classifiers that suffers from performance issues such as overfitting because of inadequate samples or missing data from the samples available while training.

The detailed structure of the proposed RBFN-EKF-AdaBoost model is shown in Figure 4-3 and the framework of the model is illustrated in Figure 4-4. As shown in Figure 4-4, it is possible to interchange the dotted part of the diagram that is the RBFN parameters optimization part of the framework with other optimization methods such as training the network with Decoupled Kalman filter or Particle Swarm Optimization (PSO) algorithms.

In the simulation, the weak classifier RBFN is fitted to a version of the dataset described in the previous section. The structure of the RBFN used is as illustrated in Figure 4-5. Then, EKF is used to train the RBFN at each iteration. The training process comprises of several training points (X_i, Y_i) where $X_i, \in X$ and $Y_i \in \{-1, +1\}$, on round $t, where\ t = 1, \ldots T$. Then compute the weighted misclassification rate of the learner and update the weighting measure used in the next round t + 1. During the training process AdaBoost called the base classifier T times, in this case 20 times. As AdaBoost trains RBF network at each round, RBFN layers are optimized using EKF to train and update the network training parameters, namely the:

standard deviation (σ), mean (μ) and the weights (w). At each training cycle, the model continuously adjusts the weights of the RBFN weak classifiers until stopping criteria is met.

During simulation process it was observed that the optimized parameters were different at the initial stage but remains unchanged after few training cycles. This indicates how quickly EKF was able to learn from the data, generate optimized RBFN training parameters. Therefore, it enables the model to converge quite fast compare to other predictive models described in the previous chapter of this thesis.

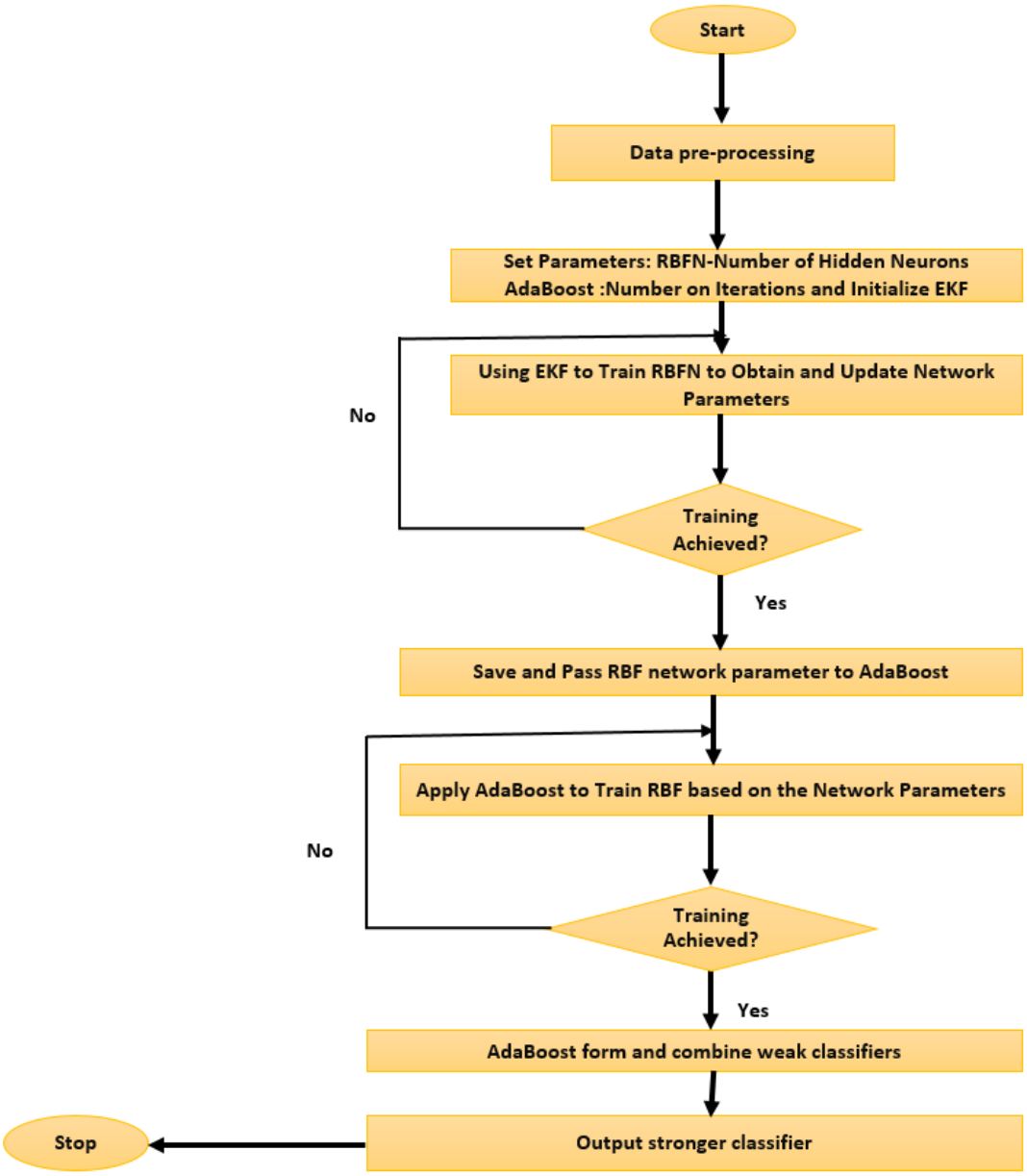


Figure 4-3: Hybrid Structure of EKF-RBFN-AdaBoost Model for Predicting Breast Cancer Survivability and Diabetic Diagnostics.

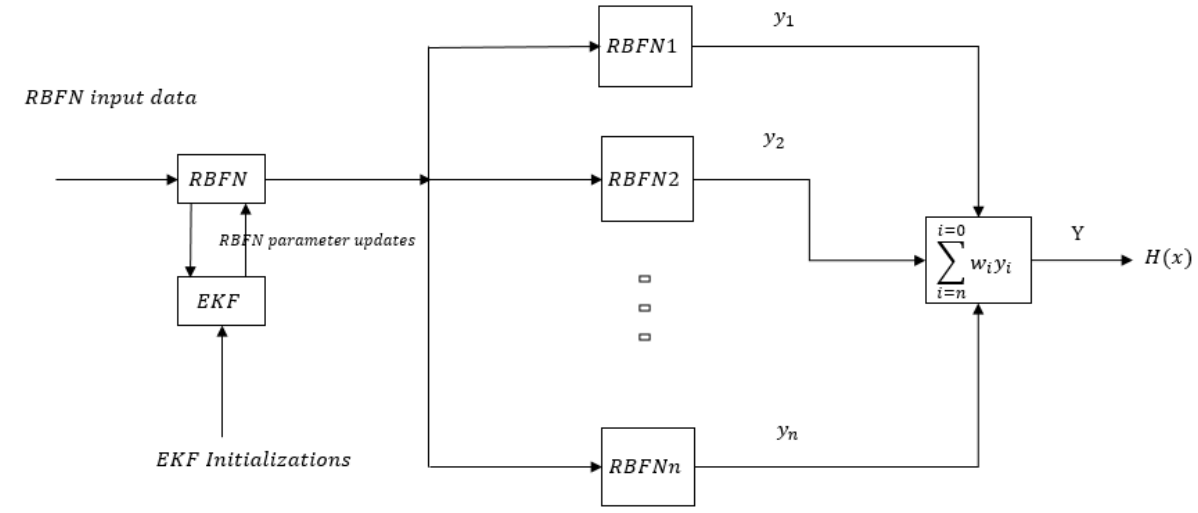


Figure 4-4: Framework of the Proposed Hybrid EKF-RBFN-AdaBoost Ensemble Model.

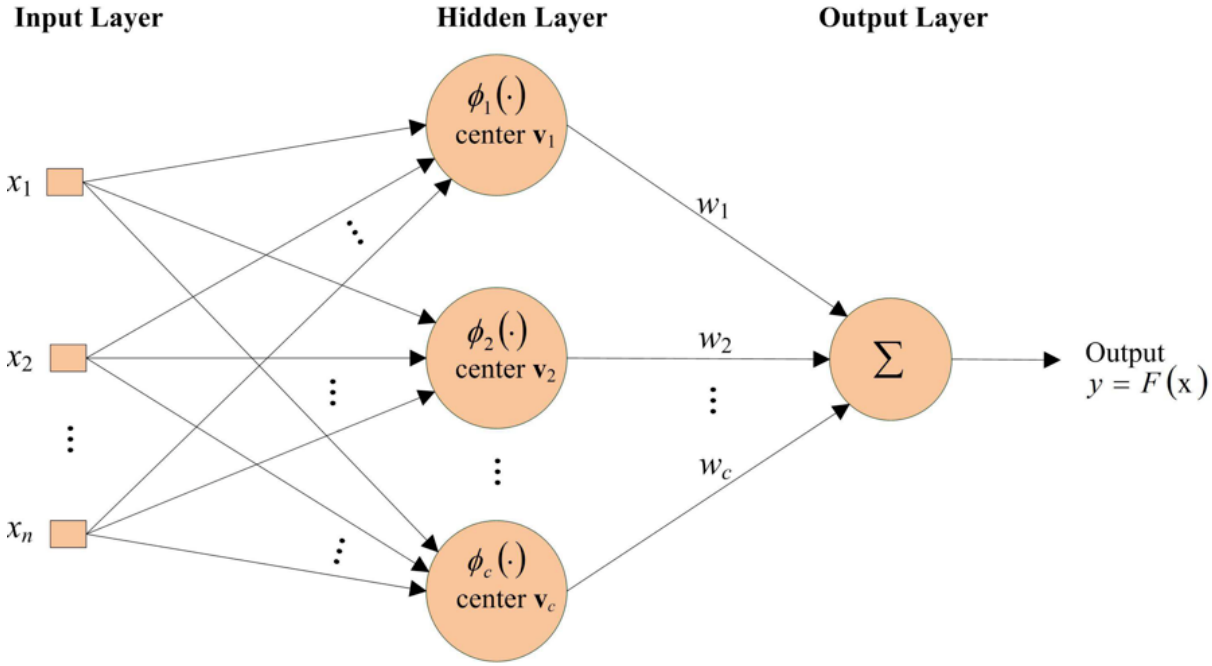


Figure 4-5: Three-layer Feed-Forward RBF Network [Adapted from *(Galar & Kumar, 2017)*].

4.6 Simulation Results and Discussion

In this section the application of EKF in training RBFN is briefly described, then training the output EKF-RBFN prototypes with AdaBoost as a meta-ensemble algorithm to obtain the final predictive output of the model.

Firstly, during the simulation process the sample is randomly divided into training and test sets, the data is normalised by replacing each feature value x by $\bar{\mathrm{x}}\ (\mathrm{x} - \mu_{\mathrm{x}})/\sigma_{\mathrm{x}}$, where μ_{x} and σ_{x} symbolise the sample mean and standard deviation respectively. Secondly, the training algorithms were initialized with prototype vectors that were randomly selected from the input data. The weight matrix w is initialized to 0, then apply EKF to train the RBFN parameters as described above. The output of the models were then passed as input to the AdaBoost model to train the prototypes and output the final predictive ensemble results as illustrated in the framework illustrated in Figure *4-4*. The model was applied to Breast Cancer Survivability, Diabetes Diagnosis datasets, Workers Absenteeism, and Clients Credit Card Defaults tasks.

The empirical performance of the proposed models on the samples compared with other benchmarked standalone and ensemble models are as illustrated in Table *4.1*, Table *4.2*, Table *4.3* and Table *4.4* respectively. The predicted results of models are also as shown in Figure 4-6 and Figure 4.7

It can be seen from the Tables and Figures that the performance of the proposed ensemble models varies, this is because different ensemble methods generated different ensemble weights and combinations of different weak classifiers on different ensemble training samples. The performance results for the proposed ensemble EKF-RBFN-AdaBoost model

based Breast Cancer survivability, Diabetes Diagnosis, Workers Absenteeism and Credit Card Defaults are as illustrated in Table *4.1*, Table *4.2*

Table *4.3* and Table *4.4* respectively. The headings of Table *4.1*, Table *4.2*

Table *4.3* and Table *4.4* are the evaluation metrics: true positive rates (TPR), false positive rates (FPR), Recall, precision, F-Measure and prediction accuracy of the model respectively.

Table 4.1: Empirical Comparison: Based on Breast Cancer Survivability Sample.

Models/Metrics	**TPR**	**FPR**	**Recall**	**Precision**	**F-Measure**	**Accuracy**
Ensemble Predictive models						
EKF-RBFN-AdaBoost (proposed model)	**0.93**	**0.03**	**0.80**	**0.97**	**0.87**	**0.96**

AdaBoostM1 with Decision stump	0.94	0.08	0.94	0.94	0.94	0.94
AdaBoostM1 with RBFN trained with K-Means	0.96	0.04	0.96	0.96	0.96	0.96
AdaBoostM1 with Random Forest	0.97	0.04	0.97	0.97	0.97	0.97
AdaBoostM1 with Support Vector Machine	0.97	0.04	0.96	0.96	0.96	0.96
Standalone predictive models						
Random Forest	0.97	0.04	0.97	0.97	0.97	0.97
Support Vector machine	0.97	0.03	0.97	0.97	0.97	0.96
K-NN	0.96	0.06	0.96	0.96	0.96	0.96
ANN	0.96	0.04	0.96	0.96	0.96	0.96
Naïve Bayes	0.96	0.03	0.96	0.97	0.96	0.96

As illustrated in Table 4.1, the simulation result shows that the proposed model outperformed other benchmark models apart from Random Forest as standalone and RF as an ensemble algorithm. Likewise, Table 4.2

Table *4.2* also shows that it was only Random Forest when simulated as an ensemble algorithm or standalone model that outperformed the proposed models by 2% respectively.

The results, as illustrated in

Table *4.3* are based on Workers Absenteeism shows the proposed model outperformed other ensemble models apart from the ensemble random forest model.

The more striking result to emerge from the result as shown in the table is that the proposed model only outperformed Support Vector machine and Naïve Bayes standalone models.

However, as shown in Table 4.4 the proposed model outperformed all other models when simulated on Clients Credit Card Defaults samples. Comparing the performances of all the models as Table 4.1, Table 4.2,

Table *4.3* and Table *4.4*. It shows that despite the topology and complexity (in terms of space and time) of ANN models, ANN models did not outperform the proposed EKF-RFFN-AdaBoost models and most of the non-complex standalone and ensemble models.

Table 4.2: Empirical Comparison Based on Diabetes Diagnosis Sample

Models/Metrics	TPR	FPR	Recall	Precision	F-Measure	Accuracy
Ensemble predictive models						
EKF-RBFN-AdaBoost (proposed model)	0.74	0.34	0.74	0.74	0.74	0.76
AdaBoostM1 with Decision stump	0.74	0.35	0.74	0.74	0.74	0.74
AdaBoostM1 with RBFN trained with K-Means	0.74	0.34	0.74	0.74	0.74	0.74
AdaBoostM1 with Random Forest	0.76	0.32	0.76	0.76	0.76	0.76
AdaBoostM1 with Support Vector Machine	0.77	0.54	0.78	0.76	0.77	0.78
Standalone predictive models						
Random Forest	0.76	0.31	0.76	0.75	0.76	0.76
Support Vector machine	0.65	0.65	0.65	0.42	0.79	0.65
K-NN	0.65	0.65	0.65	0.42	0.51	0.65
ANN	0.75	0.31	0.75	0.75	0.75	0.75
Naïve Bayes	0.76	0.31	0.76	0.76	0.76	0.76

Table 4.3: Performance Comparison Based on Smoker Absenteeism Workers

Algorithms/Measures	TPR	FPR	Recall	Precision	F-Measure	Accuracy
Predictive Models based on Ensemble Classifiers						
EKF-RBFN-AdaBoost (proposed model)	**0.95**	**0.85**	**0.95**	**0.94**	**0.95**	**96**
AdaBoostM1 + Decision stump	0.94	0.81	0.94	0.94	0.91	94
AdaBoostM1 + K-Means	0.94	0.52	0.94	0.93	0.93	94
AdaBoostM1 + with Random Forest	0.98	0.31	0.98	0.98	0.98	98
AdaBoostM1 + Support Vector Machine	0.91	0.72	0.91	0.90	0.90	92
Predictive Models Based Standalone Classifiers						
Random Forest	0.98	0.28	0.98	0.98	0.98	98
K-NN	0.98	0.52	0.94	0.93	0.93	94
Support Vector machine	0.92	0.91	0.92	0.88	0.89	92
ANN	0.97	0.34	0.97	0.96	0.97	96
Naïve Bayes	0.93	0.52	0.93	0.92	0.93	93

Table 4.4: Performance Comparison Based on Clients Credit Card Defaults

Algorithms/Measures	TPR	FPR	Recall	Precision	F-Measure	Accuracy
Predictive Models based on Ensemble Classifiers						
EKF-RBFN-AdaBoost (proposed model)	**0.80**	**0.85**	**0.82**	**0.84**	**0.88**	**85**
AdaBoostM1 with Decision stump	0.80	0.59	0.81	0.78	0.78	81
AdaBoostM1 with RBFN trained with K-Means	0.73	0.55	0.73	0.73	0.73	73
AdaBoostM1 with Random Forest	0.79	0.77	0.79	0.73	0.70	78
AdaBoostM1 with Support Vector Machine	0.78	0.54	0.78	0.76	0.76	78
Predictive Models based on Standalone Classifiers						
Random Forest	0.78	0.76	0.79	0.75	0.70	78
Support Vector machine	0.78	0.54	0.78	0.76	0.74	78
K-NN	0.73	0.55	0.73	0.73	0.73	73
ANN	-	-	-	-	-	-
Naïve Bayes	0.53	0.43	0.53	0.70	0.57	53

The comparison results of the proposed models with other benchmark ensemble models and standalone classifiers based on breast cancer, diabetes diagnostic, workers absenteeism and credit card default samples are as illustrated in Figure 4-6 and Figure 4.7 respectively.

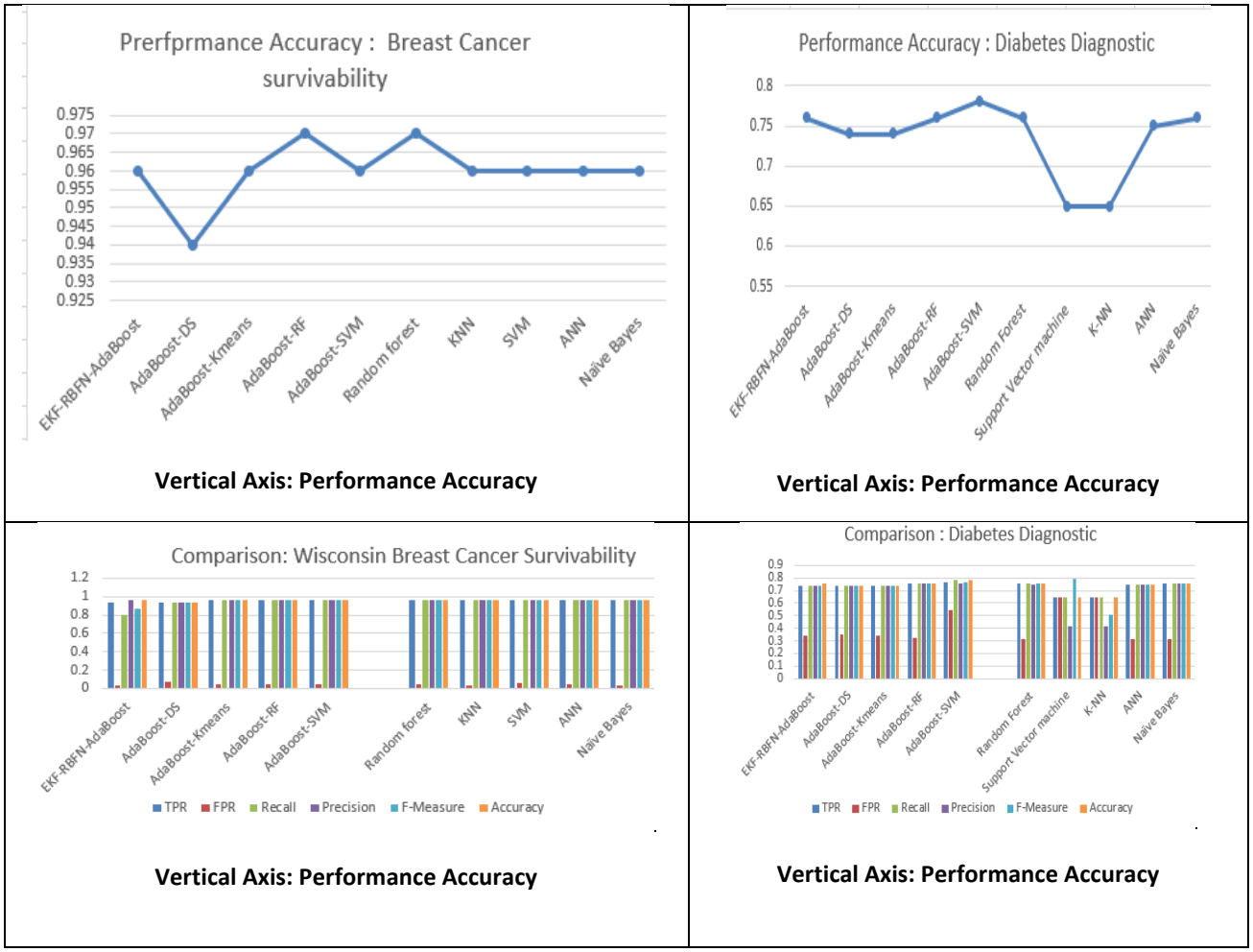


Figure 4-6: Performance of Ensemble Models on Breast Cancer and Diabetes Diagnostic Samples

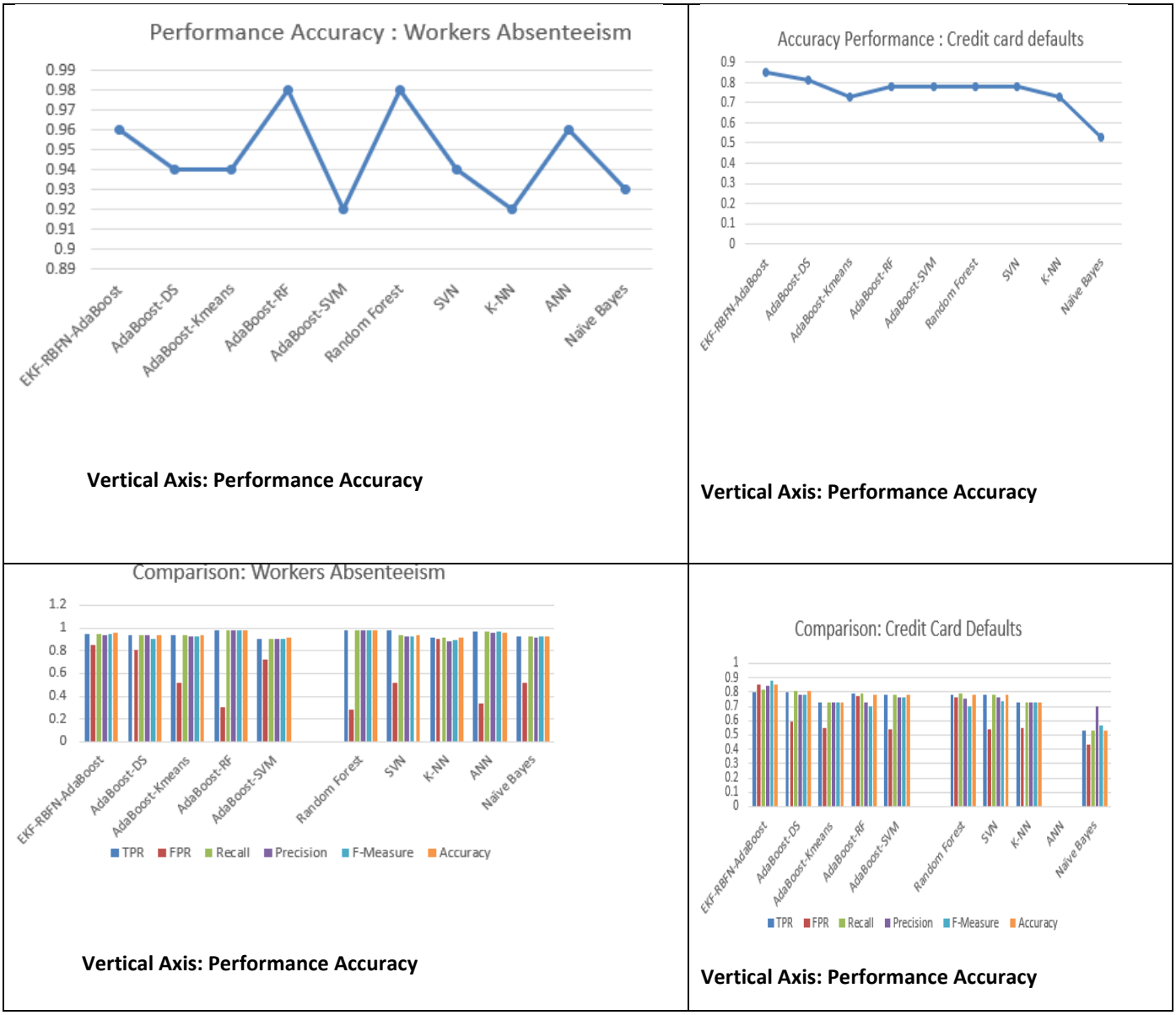


Figure 4.7 Performance of Ensemble Models: Workers absenteeism and credit card default datasets

4.7 Chapter Summary

In this chapter, a hybrid EKF-RBFN-AdaBoost ensemble model framework was developed aiming at improving prediction accuracy of ensemble models. The model uses EKF model to optimize the parameters of the RBFN members to output EKF-RBFN prototypes. It then uses AdaBoost as a meta-algorithm and the EKF-RBFN prototypes as input to train the base classifiers and output the final prediction. The framework of the proposed model could also

be applied to other optimization training methods such as Decoupled Kalman filter or Particle Swarm Optimization (PSO) algorithms.

The theoretical simulation results and the analysis of the results on breast cancer survivability, diabetic diagnosis and other databases shows that the proposed EKF-RBFN-AdaBoost model framework is a promising concept with high prediction accuracy. The model outperformed most of the established standalone and ensemble models discussed in this chapter. Additionally, the results further shows that despite the topology and complexity of the established K-NN and ANN techniques, neither of the algorithms outperforms the proposed models.

In the next chapter, the proposal of two additional ensemble methods that are based on early stopping concept and multivariable logistic model are presented respectively. The chapter also discussed the proposal of a non-invasive predictive analytical tool that is grounded on the practical application of the ensemble models proposed in this thesis.

Chapter 5: Early Stopping Approach and Multivariate Logistic Regression Models

5.1 Introduction

In this section two ensemble models namely the early stopping model and the logistic regression model are proposed. They are based on early stopping and multivariate logistic concepts, respectively. Early stopping methods are widely applied techniques to prevent poor regularization performance in gradient-based optimization problems. However, there is a limited previous research on improving ensemble classification with the AdaBoost meta-algorithm; based on the mean, standard deviation and thresholds of the training samples to estimate and build a stronger classifier.

The focus of this chapter is to address Objective 3, which is based on early stopping concept and the statistics of the training samples. To address Objective 4, which is based on development of a breast cancer multivariate logistic regression. To address Objective 5, which is based on development of a non-invasive analytical predictive tool using the predictive ensemble model developed in this thesis. To achieve this objectives the content of these session are presented in three in parts.

In the first part of this session discussion and investigation on early stopping concept (Wei, et al., 2019; Raskutti, et al., 2014) in conjunction with boosting concept to enhance predictive accuracy is presented. The proposed model extends ensemble and early stopping frameworks. In the proposed model the training of the samples are stopped early during training to avoid overfitting once there is no tangible performance improvement on the output of the training samples.

In the second part of this session, discussion and investigation on the multiple logistic regression models (Zhang, et al., 2017; Venkatesan & Sasikala, 2019) are presented. The main drive behind the use of multivariate logistic regression is to determine the significant, and the credible combination of the independent variables of training samples that best fit the dependent variable for optimum performance of predictive models. Therefore, the proposed and developed logistic model discussed in this thesis is primarily focused on feature selection of samples to enhance the accuracy performance of breast cancer survivability models.

In the third part of this session, discussion on development on non-invasive predictive analytical tools that is based on the application of ensemble predictive models proposed and developed in this work are presented.

5.2 Early Stopping Model

Early stopping or halting training after a limited number of rounds is equivalent to regularization[4] (Schapire, 2013; Schapire & Freund, 2012) to avoid incorrect prediction. In machine learning it comes to a point that the performance of an algorithm can no longer be improved due to noise and uncertainty. Recent developments in early stopping concept shows that several early stopping techniques that are based on different criteria have been suggested. Some of the significant developments among others include an approach without validation set (Maclaurin, et al., 2015), a nonparametric variation inference method (Duvenaud, et al., 2016) and a loss of validation set (Prechelt, 2012) procedure.

One of the major problems when training algorithms is the choice of the number of training epochs and when to stop training. It has been demonstrated that selection of too few epochs

[4] Regularization is a technique that protects against overfitting by constraining, smoothing, and or promoting sparsity.

during the training of samples can result in under-fitting of the models. Equally, the selection of too many epochs can lead to overfitting of the models. Many authors have demonstrated that both overfitting and under-fitting are common problems in machine learning (Zhang, et al., 2019; Anon., 2019) that generally leads to undesirable and misleading results.

In this section and other subsections the concept of early stopping, and a regularization rule that was apply to stop training of data early in order to avoid inaccurate predictive outputs as a result of overfitting or under fitting problems was discussed.

5.2.1 Model Description

The early stopping concept model proposed in this thesis is based on data dependent rules. The concept applies the mean, standard deviation, and threshold of training sample to stop further training of the model to avoid regulation error, thereby enhancing the predictive performance of the model.

The threshold function is as depict as

$$(\sigma) = \begin{cases} 1, & if\ \mu \geq 0, \\ 0, & otherwise \end{cases} \tag{5.1}$$

The model demonstrates a direct connection between the performance of early stopped iterations and the localized Gaussian complexity of the associated loss function. As argued when there is no indication for further improvement on the prediction performance there is a need to stop the training instead of going through the entire epochs that could subsequently lead to the deterioration performance of the model.

As discussed in the previous chapters, several ensemble models train weak classifiers and combine them into a more efficient single algorithm. However, there is a problem when outliers and noises are present in the training data that can lead to class mislabelling as a result under-fitting and overfitting problems.

This is because the sample weights were normalized at each iteration cycle during training such that the sum of the weights of the samples is always equal to one. Therefore, change in weights during training can be treated as a Gaussian Probability distribution.

The proposed early stopping concept minimizes overtraining and undertraining of samples by stopping further training of the samples when the stopping criterion for the classification threshold and for the early stopping criterion of the training algorithm has been reached.

The stopping criterion and the MSE is as illustrated in Equation (5.2) and Equation (5.3) respectively.

$$
\begin{aligned}
&if\ (\ (h_t(x_t) < \ (\mu_t\ + \sigma_t\ )\ \&\&\ (h_t(x_t) > \ (\mu_t - \ \sigma_t\ ))\ == true) \\
&\qquad \text{Set } h_t == 1 \text{ (classify the sample as 1)} \\
&\qquad \text{else} \\
&\qquad \text{Set } h_t == -1 \text{ (classify the sample as -1)} \\
&end\ if
\end{aligned}
\tag{5.2}
$$

$$
MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{5.3}
$$

where x_t represent training sample, μ_t is the mean and σ_t is the standard deviation of training samples. The flowchart of the proposed Early Stopping algorithm is as illustrated in Figure 5.1.

As illustrated in Figure 5-1 the classification or prediction of the sample is based on the criterion shown in Equation (5.2). In one hand, if the threshold is true the sample is classified/predicted as positive or having diabetic's ailment. On the other hand, if the threshold is false the sample is classified/predicted as negative or having no diabetics ailment.

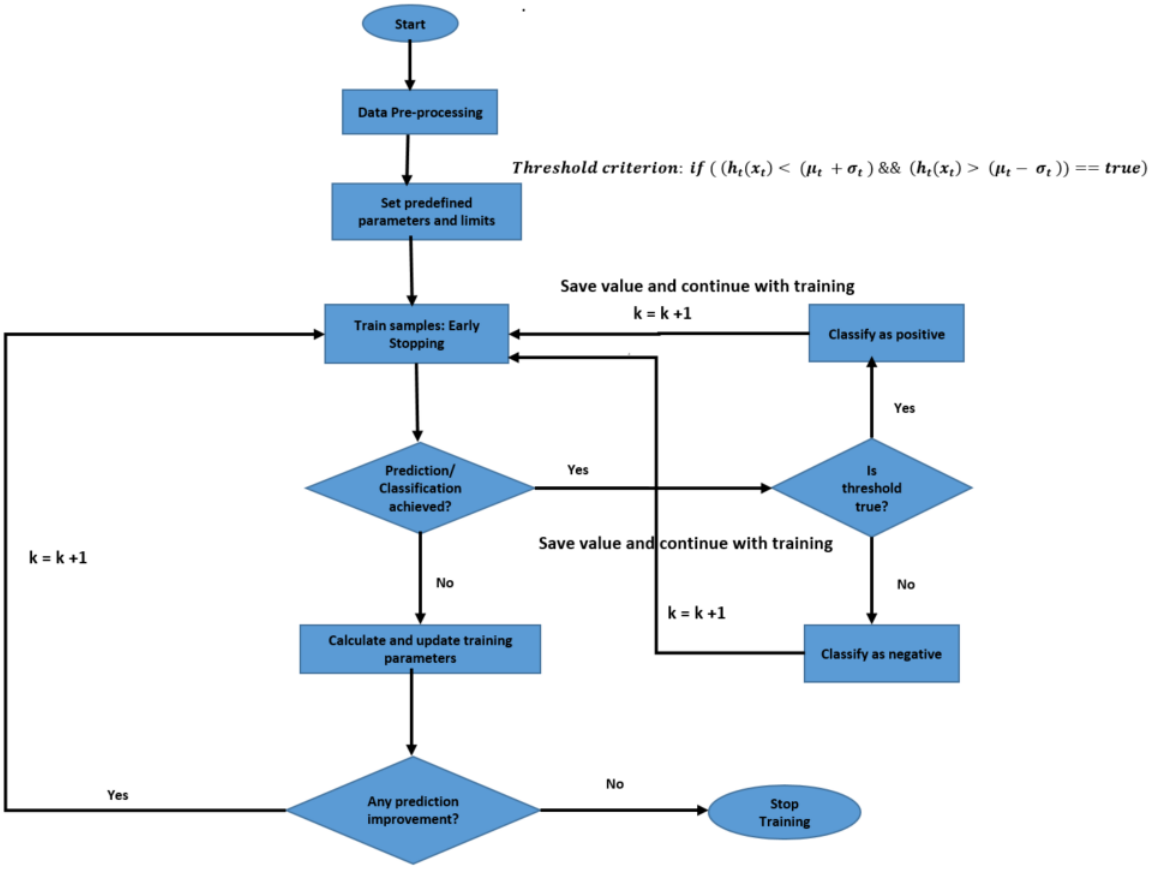


Figure 5-1 Early Stopping Algorithm Flowchart

The purpose of this approach is to avoid awarding unnecessary bias penalties to samples that are hard to train. Theoretically, as in AdaBoost as a meta-algorithm large weight penalizes hard to train weak models by increasing their weights if they predicted wrongly. Consequently, they are also compensated with little or no weights if they predicted correctly as expected at each epoch of training.

The threshold of the mean and the standard deviation of the weak classifier at each cycle of the proposed model are as expressed in Equation 5.4 and Equation 5.5, respectively, where x_t is the input data i.e., $\big(f(x_t)\big), and\ n$ is the total number of input data, μ is the mean and σ is the standard deviation of the training samples.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_t \tag{5.4}$$

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\ x_t - \mu)^2} \tag{5.5}$$

Therefore, by applying the Gaussian probability distribution, the weak classifier equation can be expressed as depict in Equation (5.4) and classification threshold as in Equation (5.5) respectively.

$$h(\boldsymbol{x} \mid \sigma^2, \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{5.6}$$

$$h_t((x_t) : X \to 1)\ s.t\ (\ (f(x_t) <\ \ (\mu_t\ + 2\sigma_t\ )\ \&\&\ \ (f(x_t) >\ \ (\mu_t - \ 2\sigma_t\ )\ ) = \\ = true \tag{5.7}$$

As demonstrated in the Equation 5.7 it shows that the classification threshold of the model was derived based on the statistical concepts of the weights *w* of the base classifier, the mean and the standard deviation of the training samples. Therefore, as the training weights *w* of the base classifier increases without performance improvement of the algorithm, the concept stop the training of the algorithm early based on the stopping rules illustrated in Equation (5.2) and Equation (5.3).

As shown in Equation (5.3), the MSE is expressed as the mean square difference between the predictive output and the expected target. In the proposed concept it is based on weight update and performance of the model after several repetitions the best selected values of μ considered during the study are 0.025, 0.01 and 0.1 respectively.

Table 5.1: Diabetics Dataset Summary

Summary	Pregnancies	Glucose	Bld_Pre	Sk_Thi	Insulin	BMI	Dia_Ped	Age
Mean	3.845	120.895	69.105	20.536	79.799	31.993	0.472	33.241
Mode	1.000	100.000	70.000	0.000	0.000	32.000	0.254	22.000
Median	3.000	117.000	72.000	23.000	30.500	32.000	0.373	29.000
Std. dev.	3.370	31.973	19.356	15.952	115.244	7.884	0.331	11.760
Minimum	0.000	0.000	0.000	0.000	0.000	0.078	0.078	21.000
Maximum	17.000	199.000	122.000	99.000	846.000	67.100	2.420	81.000

During this study, current parameters at each stage of the training were stored and updated. However, after a set of iterations and when parameter updates no longer result in prediction improvement, we stop training of the samples and use the last best parameters to output the final predictive value of the model. Therefore, it reduces the like hood of overfitting by monitoring performance of the model throughout training and restricting optimization procedure to a smaller parameter space compare to other conventional models. The summary of the diabetic dataset is illustrated in Table 5.1.

Figure 5-2 shows the density plots that illustrates the distribution of variables in the Pima-Indian-diabetics dataset; Figure 5-3 depict the scatter plots that display plot pairs of variables of the data samples that were used in training and testing the proposed early stopping model.

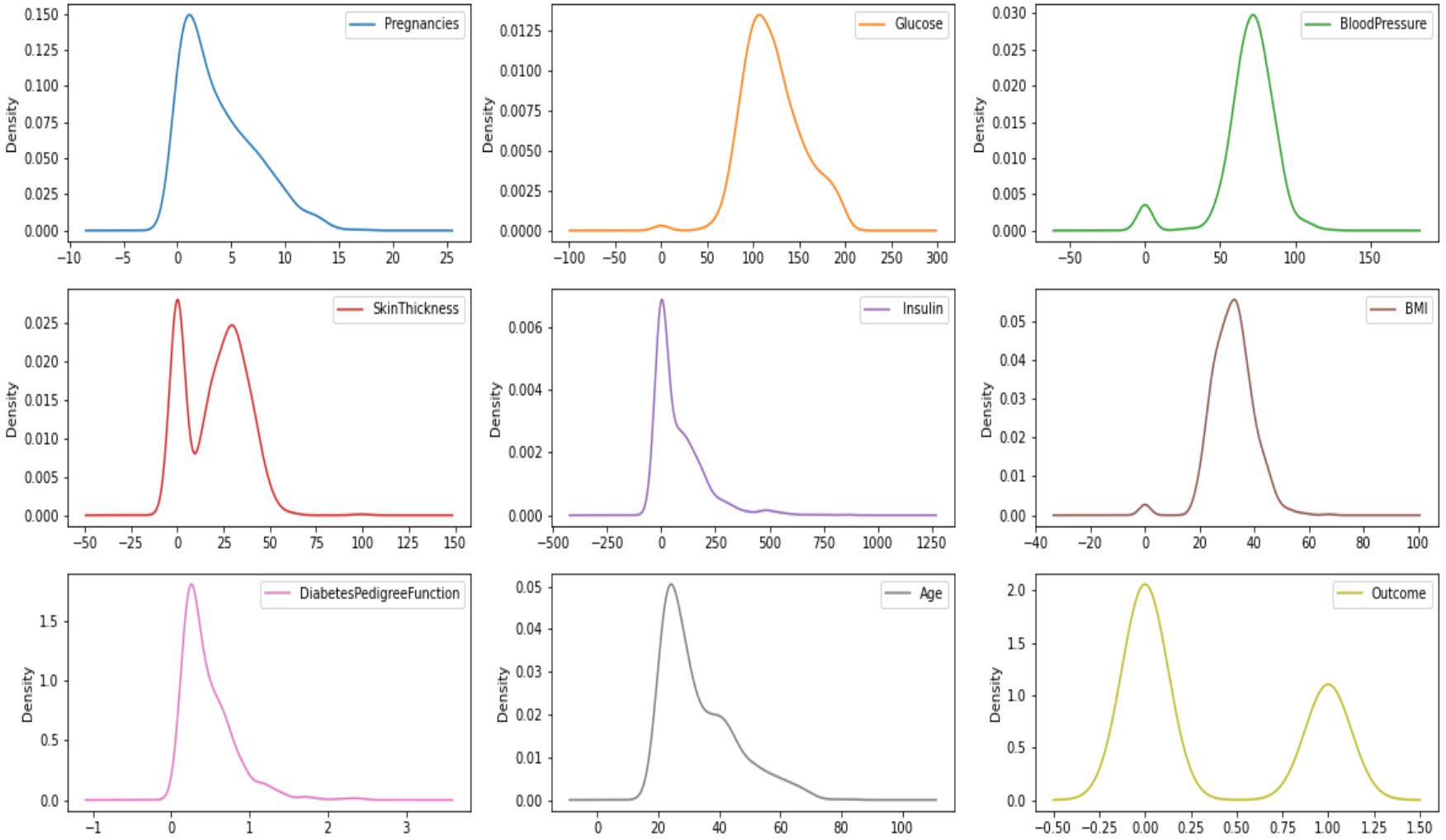


Figure 5-2: Density Plots of the Data.

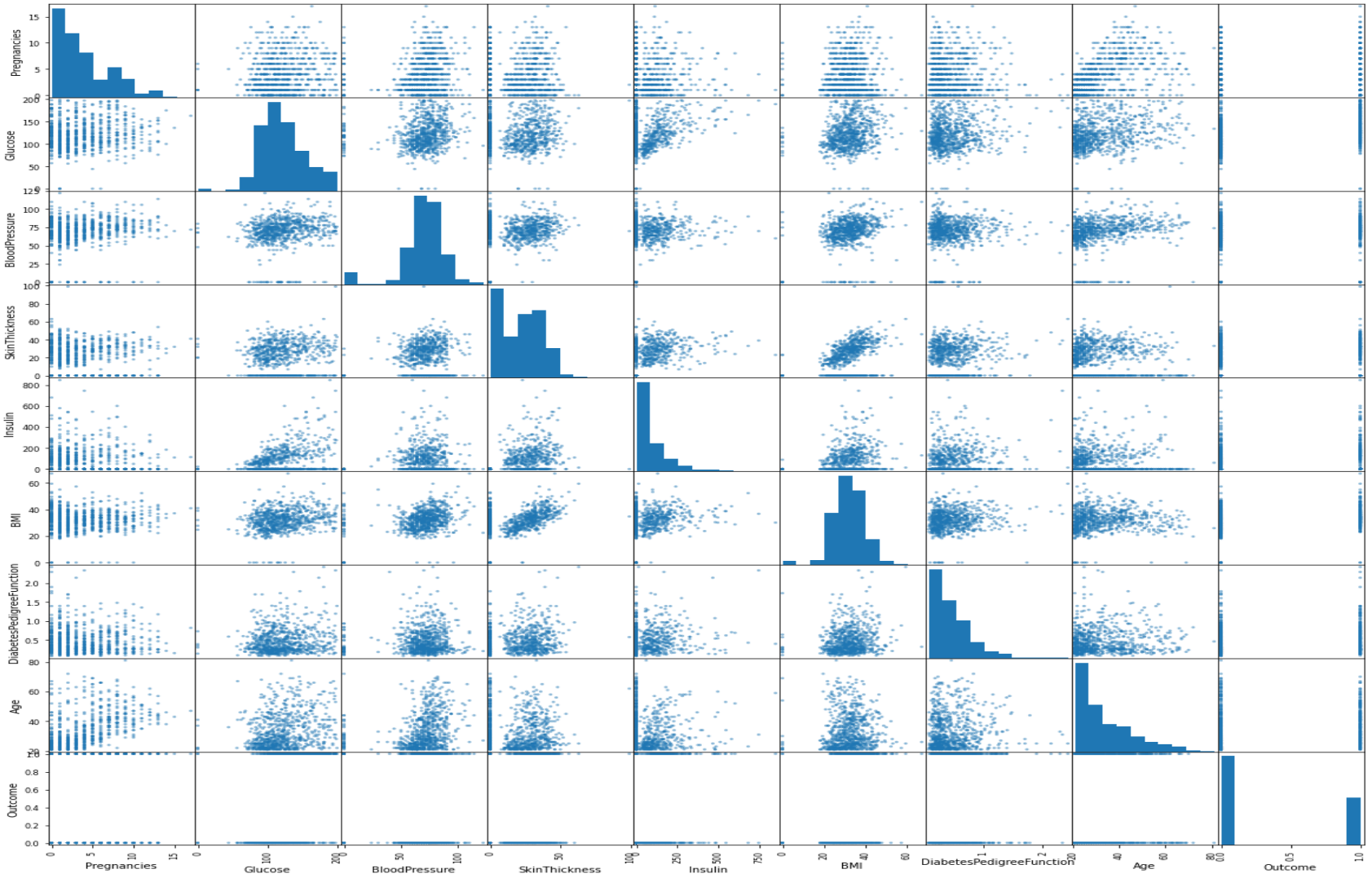


Figure 5-3: Scatter Plot Matrix of the Data

5.2.2 Simulation Description and Results

In this section the application of the proposed Early Stopping concept and simulation results to in developing a model named London Southbank Boost loss (LS-Boost loss) are briefly described. During the simulation process the sample is randomly divided into training, test and validation sets. The data is normalised by replacing each feature value x by $\bar{x}\,(x - \mu_x)/\sigma_x$, where μ_x and σ_x symbolise the sample mean and standard deviation respectively. The proposed model was applied to Pima-Indian-diabetics dataset with eight features and 768 instances. The dataset is used to predict whether or not a patient has diabetes that is centred on diagnostic features that are included in the dataset.

After the model was developed it was applied to estimate prediction performance of the Pima-Indian-diabetics datasets as illustrated above. Figure 5-4 and Figure 5-5 shows that the learning curve and the LS-Boost loss for each epoch of the proposed models on the diabetics samples. The performance results of the model are compared with other stopping algorithms as illustrated in Table 5.2.

The analysis in Table 5.2, shows that the early stopping method of the proposed model stops training earliest, with lowest number of epochs compare with other models. However, it has a lower precision performance compare with LOGLOSS and MAE but same as AUC and RMSLE early stopping algorithms. Albeit, the proposed method demonstrates to some extent that the overfitting problem in sample can be reduced and therefore the generalization ability of ensemble algorithm can be improved. The training and testing curves of the proposed model is as shown in Figure 5-4 and Figure 5-5 respectively. The vertical axis depict the classification error and the horizontal axis represents the number of iterations in both figures. As can be

seen in the training and testing curves in Figure 5-4 and Figure 5-5 respectively; the training of samples were stopped when the validation error began to rise without significant performance improvement of the model.

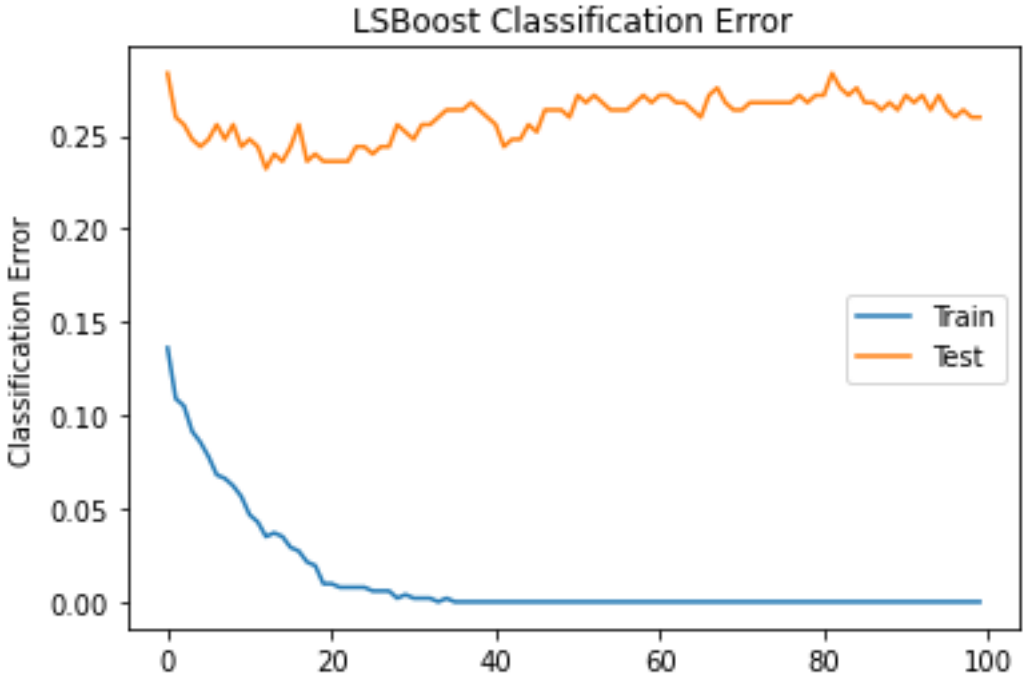


Figure 5-4: LS-Boost Learning Curve

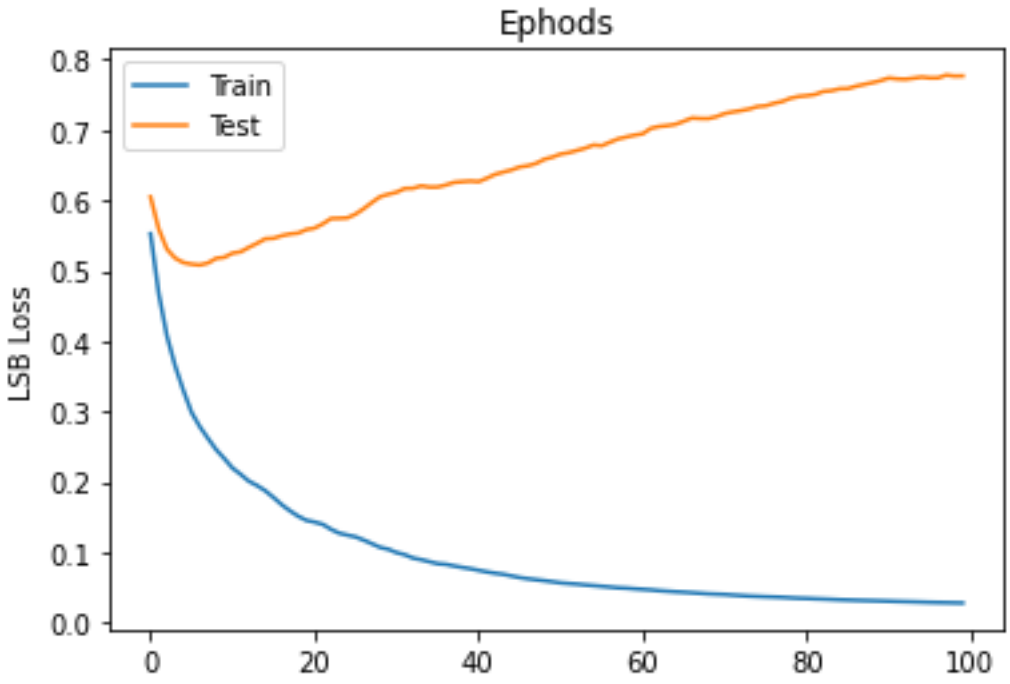


Figure 5-5: LS-Boost Loss for each Epoch based on Pima-Indian-diabetics samples

Table 5.2: Comparion of LS-Boost Results with other Early Stopping models

Stopping Algorithms	**Stopping Error**	***Number of Epochs***	***Training performance***	***Validation performance***
LS-Boost	**0.42344**	**15**	**74.02**	**74.41**
Log-Loss	0.48727	32	77.95	75.00
MAE	0.29870	31	74.02	76.38
AUC	0.25984	16	74.02	74.41
RMSLE	0.28496	16	74.02	74.41

5.3 Multivariable Logistic Regression Model

In recent years, there has been an increasing interest in the application of multivariate logistic regression models for the analysis and predictions in different fields (Sarker, 2021; Lynam, et al., 2020; Mai, et al., 2019; Joshi & Dhakai, 2021) and professions.

The proposed model in this session is similar to the linear regression technique to predict the dependent variable of samples. The models utilizes logistic regression models and Stepwise selection algorithm in R programming in understanding and predicting breast cancer survivability and diabetics' diagnostic of patents. Though, unlike the former the only exception is that the outcome of the logistic regression model is one dichotomous. This is analogous to the linear regression, it uses an equation as representation to predict the output of a categorical dependable variable that is as either: success or failure; yes or no;

infected or not infected, etc. However, far too little attention has been directed toward the application of the concept in addressing the prediction of invasive ailments such as diabetics diagnostic and breast cancer survivability. Despite its long predictive success, multivariable logistic regression concept suffers from several major drawbacks and controversial issues (McDonald, 2014) such as the assumption of linearity between the dependent variable and the independent variables.

Another problem is the multi-collinearity issue that exist when an independent variable is highly correlated with one or more of the other independent variables in a multiple regression equation. This is because it undermines the statistical significance of an independent variable (Lin, 2008; Allen, 1997) and the sparsity of the data.

In addressing this issue, using Surveillance Epidemiology and End Results (SEER) breast cancer dataset, Chen, *et. al.* applied a learning-based approach to high dimensional, high-volume, and high-sparsity data to identify critical casual attributions that might affect the survival of a breast cancer patient (Chen, et al., 2021).

However, the multivariable logistic regression model projected in this section is based on a single dichotomous outcome taking into consideration t critical casual attributions that might affect the breast cancer survivability. The model is grounded on the multiple independent variables to improve predictive output of training samples.

5.3.1 Simulation Description and Results

The proposed multivariable logistic regression and the simulation results of the proposed model are briefly described in this section. The main drive behind the use of multivariable logistic regression is to determine the significant and the credible combination of the independent variables that best fit the dependent variable. The multivariable logistic

regression model was constructed by modelling various independent variables of breast cancer survivability sample. The dataset has eight features and 698 instances. The sample was normalised by replacing each feature value x by $\bar{x}\ (x - \mu_x)/\sigma_x$, where μ_x and σ_x symbolise the sample mean and standard deviation respectively. The sample used for the simulation was randomly divided into the training, test and validation sets. The summary of the dataset is illustrated in Table 5.3. It shows the statistical valuation metrics, namely mean, mode, median, standard deviation values and the minimum and maximum values of the samples(as illustrated in the footnote) used in the study.

The dataset is used to predict whether or not a patient will survive breast cancer that is centred on diagnostic features that are included the survivability dataset. Thus it represents a multivariable regression model which clinicians can easily use to predict the likelihood of breast cancer survivability in their patients. Figure 5-6 shows the density plots that illustrates the distribution of variables in the breast cancer survivability dataset. Figure 5-7 depict the scatter plots that display plot pairs of variables of the data samples that were used in training and testing the proposed multivariate logistic model.

Table 5.3: Cancer Dataset Summary[5]

Summary	C_Thick	C_Size	C_Sha	Mar_Adh	S_Cs	Ba_Nu	Bl_Ch	N_Nu	Mit
Mean	4.418	3.134	3.207	2.807	3.216	3.545	3.438	2.867	1.589
Mode	1.000	1.000	1.000	1.000	1.000	2.000	2.000	1.000	1.000
Median	4.000	1.000	1.000	1.000	2.000	1.000	3.000	1.000	1.000
Std. dev.	2.816	3.051	2.972	2.855	2.214	3.644	2.438	3.054	1.715
Minimum	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Maximum	10.000	10.000	10.000	10.000	10.000	10.000	10.000	10.000	10.000

Stepwise selection in R was used for the variable selection of the samples. It is a combination of forward and backward selection that sequentially add the most contributed predictors, like the forward selection model. Like the backward model it also remove variables that no longer contributes improvement in fitting the model *(Zhang & Z., 2017; James, et al., 2014)*. The selected features are as illustrated in Table 5.4 while the statistical significance of the selected features is as shown in Table 5.5.

[5] In all cases N=699
C_Thick = Clump Thickness; C_Size= Cell_Size_Uniformity;C_Sha= Cell Shape Uniformity; Mar_Adh= Marginal Adhesion; S_CS= Single_Epi_Cell_Size; Ba_Nu= Bare Nuclei; Bl_Ch = Bland Chromatin; N_Nu = Normal Nucleoli; Mit= Mitoses

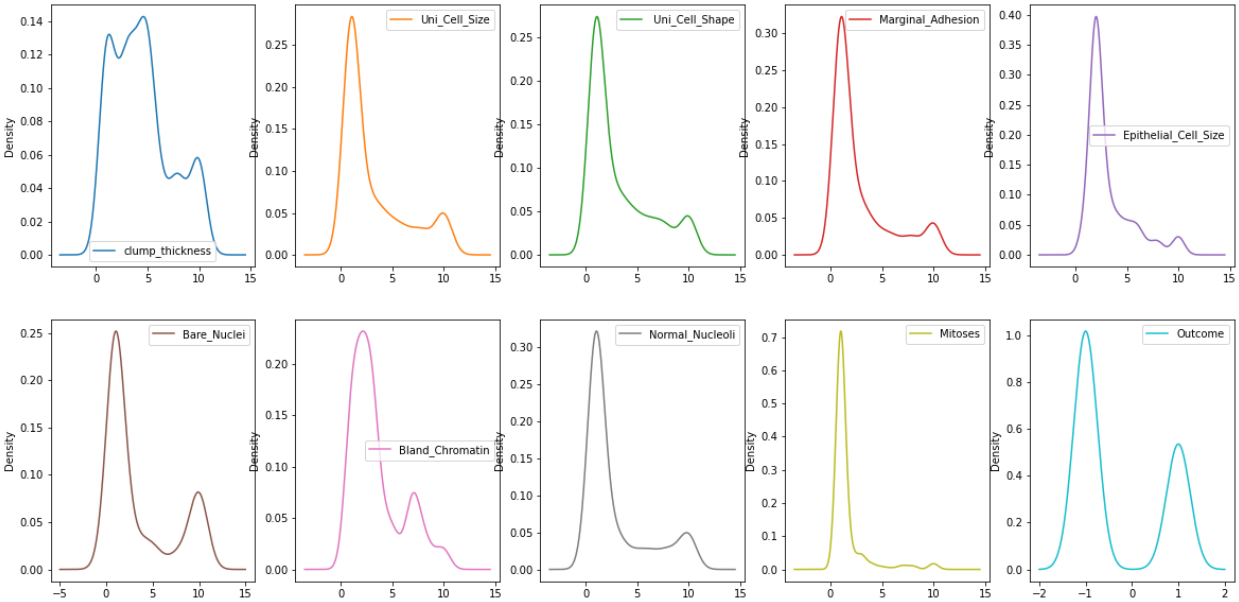


Figure 5-6: Density Plots of the Data: Breast Cancer Survivability.

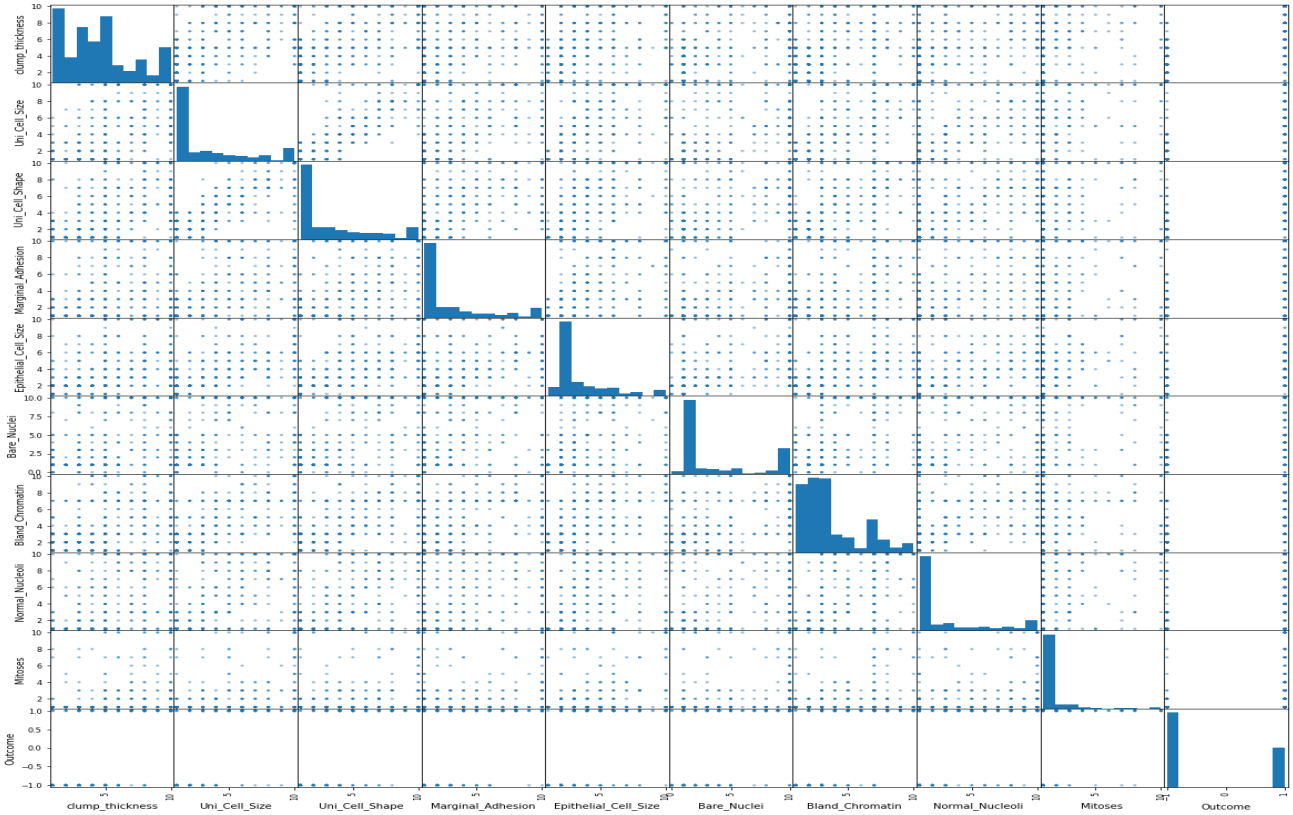


Figure 5-7 Matrix Scatter Plot of the Data: Breast Cancer Survivability

The evaluated details of the selected features, the correlations and some of the statistical properties of the model is as illustrated in Table 5.4 and in Table 5.5 respectively. These are used in formulating the multivariable logistic regression equation, as illustrated in Equation (5.8).

$$ln\,(Y) = -10.10394 + 0.53501 * X1 - 0.00628 * X2 + 0.32271 * X3 + 0.33064 * X4 + 0.09663 * X5 + 0.38303 * X6 + 0.44719 * X7 + 0.21303 * X8 + 0.53484 * X9 \quad (5.8)$$

Table 5.4: Multivariate Logistic Regression Model to Predict Breast Cancer

	Intercept	Std. Error	Z-value	Significant
Variables	-10.10394	1.17488	-8.600	0
Clump thickness (X1)	0.53501	0.14202	3.767	0
Uni-Cell Size (X2)	-0.00628	0.20908	-0.030	1
Uni-Cell Shape (X3)	0.32271	0.23060	1.399	1
Single epithelia (X4)	20.33064	0.12345	2.678	0.001
Epithelial Cell size (X5)	0.09663	0.15659	0.617	1
Bare Nuclei (X6)	0.38303	0.09384	4.082	0
Bland Chromatin (x7)	0.44719	0.17138	2.609	0.001
Normal Nucleoli (X8)	0.21303	0.11287	1.887	0.1
Mitoses (X9)	0.53484	0.32877	1.627	1

Some of the statistical performance of the proposed multivariate logistic model are as shown in Table 5.6 and Figure 5-8. As can be seen in Table 5.6, the significant level of accuracy among

the dependent features of the breast cancer samples ranges from 79% to 92%. This demonstrates that all the variables are tightly associated. Therefore they contributes significantly to the overall prediction accuracy of the proposed logistic model.

Table 5.5: Statistical Significance of Cancer-Dataset Parameters used in Predicting Cancer Prognosis

Metrics/ Features	Accuracy	RSME	KAPPA	TP	FP	Precision	Recall	F-Measure	Features
Clump thickness	85	0.324	0.651	0.855	0.260	0.874	0.989	0.899	X1
Uni-Cell Size	92	0.240	0.823	0.919	0.076	0.923	0.919	0.920	X2
Uni-Cell Shape	92	0.234	0.826	0.920	0.076	0.923	0.919	0.919	X3
Single epithelia	85	0.326	0.673	0.859	0.222	0.863	0.859	0.854	X4
Epithelial Cell Size	90	0.290	0.786	9.900	0.096	0.905	0.900	0.901	X5
Bare Nuclei	90	0.269	0.798	0.908	0.106	0.908	0.908	0.908	X6
Bland Chromatin	90	0.270	0.800	0.908	0.106	0.908	0.908	0.908	X7
Normal Nucleoli	89	0.302	0.769	0.898	0.146	0.897	0.898	0.896	X8
Mitoses	79	0.406	0.473	0.788	0.369	0.810	0.788	0.765	X9

During the training and simulation stages it was observed that the classification errors of the model is directly influenced by the percentage of dataset used in training, and the percentage of samples used in testing the model. In this particular case, as the percentage of training data increases the misclassification error also increases until it got to a pick at 80% of training data before it starts to drop. The performance of the model using different training

and testing samples, and the corresponding performance of the model are as shown in Table 5.6. The performance correlation of the misclassification error and the corresponding percentage of the training data is as illustrated in Figure 5-8. This could be as result of the features of the proposed learning algorithm and training samples (Schmidt, et al., 2019; RaEng., 2014) used during the research.

Table 5.6: Training Size and Corresponding Classification Error

Classification Error	Training data (%)	Testing Data (%)
3.41	40	60
3.72	45	55
4.09	50	50
4.55	55	45
4.74	60	40
4.42	65	35
6.34	70	30
6.43	75	25
6.57	80	20
4.85	85	15
4.35	90	10

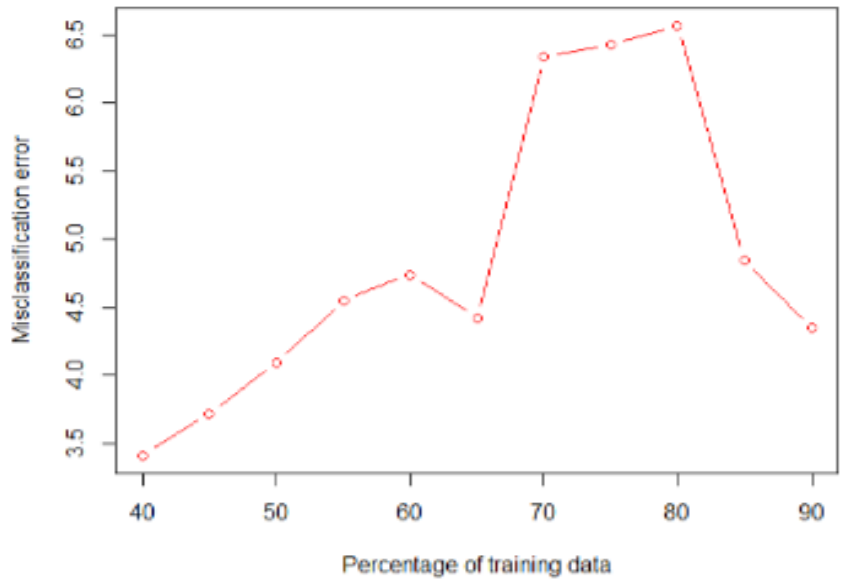


Figure 5-8 Percentage Plot of Training Data and Classification Error.

5.4 Non-Invasive Predictive Analytic Tools

Predictive analytics is both existing and an expansive area of research. The concept involves combination of statistical methods in data mining, predictive modelling, and machine learning that analyse historical data to make predictions about future events (Nyce, 2007). It is widely used to boost marketing results, machine diagnosis and automated processes in the industrial sectors. Recently, it has also been used extensively in the healthcare sector to detect early signs of patients' deterioration in the hospitals; identifying at-risk patients in the comfort of their homes to prevent hospital readmissions (Philips., 2020), thus saving taxpayer's money, time and lives of patients.

According to the WHO the overall number of people diagnosed with cancer has nearly doubled, from an estimated 10 million in 2000 to 19.3 million in 2020. It has overtaken lung cancer as the world's mostly commonly diagnosed cancer with 19.3 million new cases and 10.0 million deaths in 2020 (WHOa, 2021). Similarly, about 422 million people worldwide have

diabetes, and 1.6 million deaths are directly attributed to diabetes each year. Report shows that the number of cases and the prevalence of diabetes have been steadily increasing in the last decades (WHOb, 2020).

Similarly, about 422 million people worldwide have diabetes, and 1.6 million deaths are directly attributed to diabetes each year. Report further shows that the number of cases and the prevalence of diabetes have been steadily increasing in the last decades (WHOb, 2020).

Technology is changing and ever evolving, therefore the quest for this phase of the thesis is to develop non-invasive predictive analytic tools for the prediction of breast cancer and diabetics diagnostics in patients.

Therefore, the main drive of this model is to exploit the potentials of the proposed extended ensemble models and utilize it in the medical field as a heath software tool in predicting breast cancer survivability of patients and diabetic diagnosis without risks or and privacy invasion. This is achieved by extending the concepts of some of the proposed models in this study as an easy to use predictive analytical tool by healthcare practitioners. A number of web technology tools and programming languages were selected for the development and testing of the analytic tool.

5.4.1 Predictive Analytic Tools

Predictive analytic tools are crucial to identify patterns and forecast the likelihood of future outcomes based on historical data. It has been applied in various fields and industries: finance, healthcare, transportation, business, quality improvement and fraud prevention. A number of predictive breast cancer models have been proposed (NIH, 2020; Colozza, et al., 2005; Singh & Thakral, 2018) using different set of risk input and output based factors. Similarly several diagnostic model tools for predicting diabetes types 1 and 2 have been

proposed using different classifiers (Li, et al., 2021; Yang, et al., 2021; Saiti, et al., 2020) and several other methods. However, many of these models are costly, privacy invasive and are mostly designed for health-care professionals use only.

5.4.2 System design: Breast Cancer and Diabetic Diagnostic Tools

The analytical tool proposed in this section is privacy protected and non-invasive, cost effective and could be used at any time through the World Wide Web portals. The proposed breast cancer analytical and diabetic analytical tool described in this section includes a separate digital interfaces where users can supply breast cancer or diabetic diagnostic data.

The proposed predictive tool has two user interfaces: one for the breast cancer prediction and the second interface for the prediction of diabetes diagnostic disease. Based on the selected interface by the user, the required predictive model interface is loaded, the tool displays a form for user to enter the required features of the selected predictive model.

The tool validates the parameters that were entered by the user, otherwise appropriate messages for correction are displayed. When the user submits the form content; the tool calls a back end function that integrates the entered data with the selected predictive model proposed and developed in this study. This could be either predictive cancer model or diabetic predictive model depending on the selected technique at the interface level.

At the backend, the tool maps the submitted data with functions that integrates appropriate predictive model in decision-making. Based on the validated data the tool make appropriate decision. The decision is send back to the client interface that displays appropriate predictive message to the user. The user interface flowchart of the proposed non-invasive analytical tool is as illustrated in Figure 5-9. The software lifecycle development cycle (SDLC) methodology (Langer, 2016) was used to develop, build and test the proposed system.

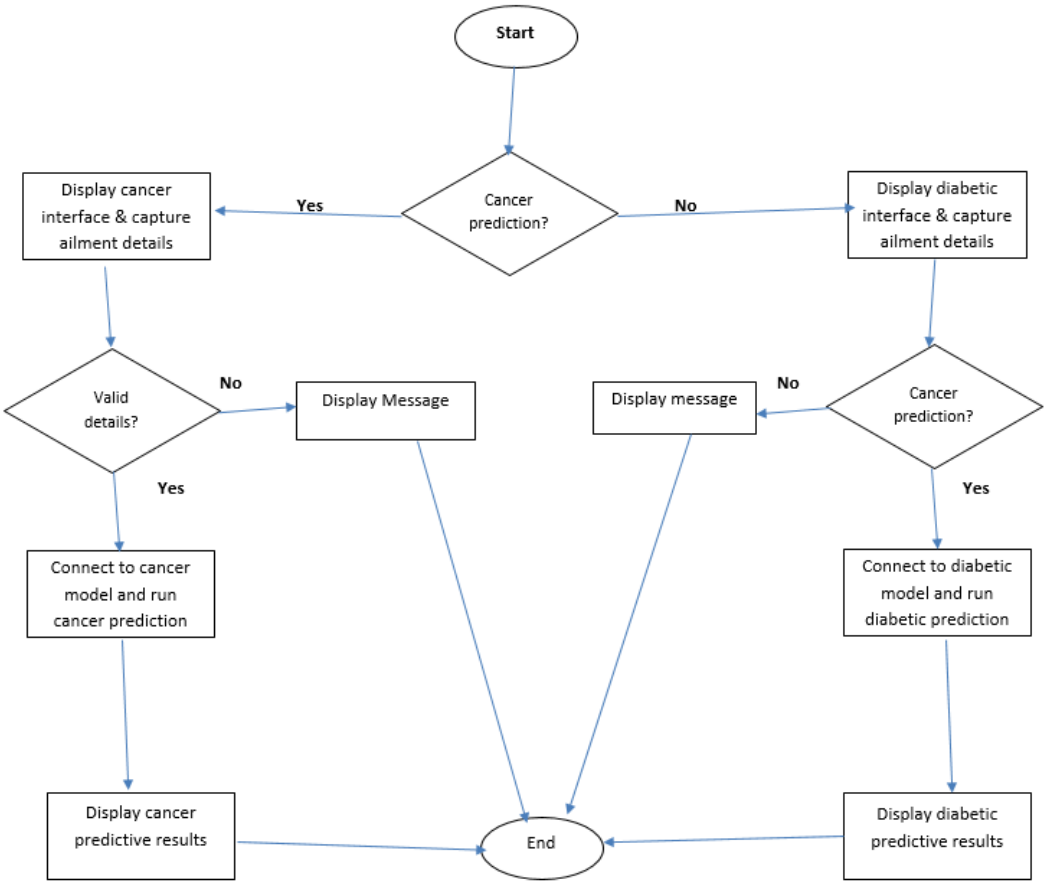


Figure 5-9: Simplified Guidelines for the Selection of Appropriate Predictive Model

The Use Case diagram of the proposed web application analytic tool is as depict in Figure 5-10: Use Case for Predicting Breast Cancer Survivability and Diabetes' Diagnostics. It has two different types of users:

I. Administrator and support staff – These are users with high level of access to the software.

II. Patients – These are breast cancer and diabetics diagnostic patients with read only access.

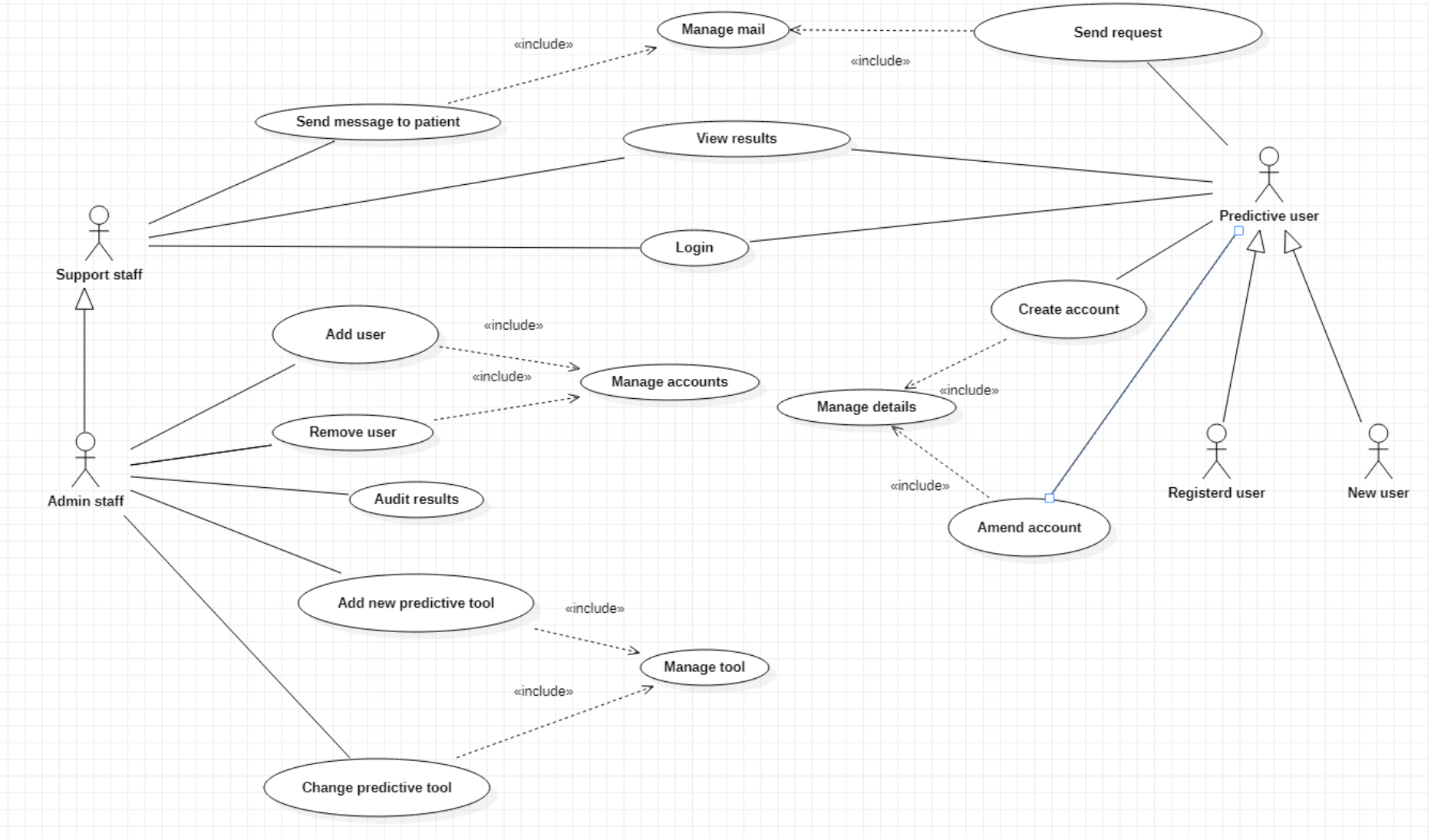


Figure 5-10: Use Case for Predicting Breast Cancer Survivability and Diabetes' Diagnostics

5.4.3 Simulation Results and Discussion

Some of the predicted results of the proposed tool based on breast cancer samples is as illustrated in Table 5.7. The table displays the predictive parameters of the testing sample (and their corresponding values as descried in the footnotes), the expected label values and the predicted label values. During the testing phase several tests were carried out in order to validate the performance efficiency of the model. As illustrated in Table 5.7, it is apparent from the table performance average of the model for breast cancer is 100% after 10 randomly tests were carried out.

Table 5.7: Some Results of the Evaluation Tool Using Breast Cancer Samples[6]

Regression variables/ Test number	***V1***	***V2***	***V3***	***V4***	***V5***	***V6***	***V7***	***V8***	***V9***	***Expected Label***	***Predicted Label***
1	1	1	1	1	2	1	1	1	1	-1	-1
2	5	8	9	4	3	10	7	1	1	1	1
3	4	1	1	1	1	1	2	1	1	-1	-1
3	5	10	10	10	6	10	6	5	2	1	1
5	5	1	2	10	4	5	2	1	1	-1	-1
6	3	1	1	1	1	1	2	1	1	-1	-1
7	1	1	1	1	1	1	1	1	1	-1	-1
8	4	2	1	1	2	1	1	1	1	-1	-1
9	3	3	1	1	2	1	1	1	1	-1	-1
10	6	6	7	10	3	10	8	10	2	1	1
Breast Cancer Predictive Model Average Performance											***100%***

[6] V1, V2…V9 are the Breast Cancer predictive model parameters. These are Clump thickness, Uni-Cell Size, Uni-Cell Shape, Single epithelia, Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli, and Mitoses respectively

Likewise, some of the predicted results of the proposed tool based on Diabetes Diagnosis Samples is as presented in Table 5.8.

As illustrated in Table 5.8, it shows the predictive parameters of the sample (with their equivalent values as descried in the footnotes), the expected label values and the predicted label values. Several tests were also carried to ascertain the performance accuracy of the proposed analytical non-invasive model. Table 5.8 depicts the performance outcome of the model based on 10 randomly test samples. The results, as shown Table 5.8, indicates that the average performance of the model on Diabetes Diagnosis Samples was 90%.

Table 5.8: Some Results of the Evaluation Tool Using Diabetes Diagnosis Samples[7]

Test variables/ Test number	***V1***	***V2***	***V3***	***V4***	***V5***	***V6***	***V7***	***V8***	***Expected Class Label***	***Predicted Class Label***
1	1	109	38	18	120	23.1	0.407	26	*-1*	*-1*
2	3	129	92	49	155	36.4	0.968	32	*1*	*1*
3	8	100	74	40	215	39.4	0.661	43	*1*	*1*
3	3	128	72	25	190	32.4	0.549	27	*1*	*-1*
5	2	197	70	99	0	34.7	0.575	62	*1*	*1*
6	0	151	90	46	0	42.1	0.371	21	*1*	*1*
7	6	109	60	27	0	25	0.206	27	*-1*	*-1*
8	12	121	78	17	0	26.5	0.259	62	*-1*	*-1*
9	1	124	60	32	0	35.8	0.514	21	*-1*	*-1*
10	2	68	62	13	15	20.1	0.257	23	*-1*	*-1*
Diabetics Diagnostic Predictive Model Average Performance										***90%***

[7] V1, V2…V8 are the Diabetic diagnostic predictive model. These are Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, diabetes Pedigree Function and Age respectively.

The results as shown Table 5.8 indicates that only 1 in 10 of the selected samples is incorrectly classified and 9 in 10 selected samples are correctly classified. Therefore, the prediction of the errors model are moderately low compared with other predictive analytical models.

5.5 Chapter Summary

In this chapter, early stopping forecasting model has been developed to prevent poor regularization performance problems in ensemble model, it was applied to the Pima Indians Diabetic and breast cancer samples.

A multivariate logistic regression models was also developed in order to improve the predictive performance of breast cancer model. The model was applied to Breast Cancer survivability and diabetic diagnostic tasks. Both models show promising results that could be effectively used in predicting diabetic diagnostic and breast cancer survivability in patients respectively.

This chapter also focus on practical application of the predictive models proposed in this thesis by proposing a non-invasive analytical breast cancer survivability and diabetic diagnostic tools. That could be used in predicting breast cancer survivability and diabetic diagnostic. The analytical breast cancer and the diabetic diagnostic tools integrates the breast cancer predictive model and diabetic diagnostic model that were developed in this thesis respectively. Unlike most of the existing breast cancer and diabetic predictive methods, the

tools are cost effective, it does not intrude privacy of users and can be used online. The predictive performance efficiency of the tools are 100% and 90% on randomly selected tests using breast cancer survivability and diabetic diagnostic samples respectively. Overall, the results in this chapter encouraging. The concepts are simple and non-invasive. The next chapter presents the discussion of this thesis.

Chapter 6: Discussion

6.1 Introduction

The previous chapter presented the performances and the findings obtained from the proposed early stopping, multivariate predictive concepts and the non-invasive analytic tools that were developed and tested.

The purpose of this chapter is to address Objectives 6 and 7, which are based on discussion and reflection of the research investigations, findings, and analysing the performance outputs of the proposed research models. In addition, to clarify on some of the research's benefits and limitations, and to make recommendations for further research respectively.

After literature review of previous work on ensemble modelling and establishing some research gaps, research hypothesis were raised. Then experimental investigations on the performances of AdaBoost implementation methods, standalone and ensemble techniques were also conducted. Doing this offers avenue to provide answers to the research hypothesis and address some of the misunderstanding concepts among researchers in the area of ensemble development. This further provides adequate knowledge on how to investigate, propose and develop new standalone and ensemble methods in order to enhance their predictive performances and reliabilities.

The initial investigations specifically covered during phase one of this study are:

I. The implementation and comparisons of AdaBoost implementation methods namely: implementation by resampling, implementation by reweighing and their effects on ensemble modelling performance.

II. The development and performance comparison of standalone, complex (in terms of space and time) and non- complex ensemble models.

During this study in other phases of the thesis the following new ensemble predictive models and analytical tools were specifically proposed, developed and tested after the initial investigation:

I. **An ensemble EKF-RBFN-AdaBoost model** - In phase two of the research a model that uses EKF to train and enhance the training parameters of RBNF models was proposed and developed. Then ensemble AdaBoost, a meta-algorithm was applied to train the EKF-RBFN prototypes as a committee of classifiers, and combine them to obtain a better classifier as a predictive output. To demonstrate the effective and accuracy of the EKF-RBFN-AdaBoost model concept, it was applied to Breast Cancer samples and several other samples.

II. **Early Stopping Approach Model** – To improve the performance of ensemble model, a concept based on the early stopping idea and statistical performance of the training weak learners was proposed. The purpose of this was to avoid overfitting ensemble models once certain criteria and threshold were reached. To demonstrate the effective and accuracy of the concept it was applied to diabetics' diagnostic samples.

III. **Multivariate Logistic Regression Models** – To minimize the complexity of predictive models, a model based on the concept of multivariate logistic regression was proposed. The model uses a multivariate equation to represent and predict the output of a categorical dependable variable. To demonstrate the effective and accuracy of the regression model it was applied to breast cancer survivability modelling tasks.

IV. **Non-invasive Predictive Analytical Tools** – To further integrate the theoretical aspect of the models developed in this thesis with practical applications; ensemble non-invasive predictive analytical tools based on the proposed ensemble models in this thesis were implemented. The proposed models in this thesis served as backend predictive objects for the proposed non-invasive analytic tools. To demonstrate the effective and accuracy of the tools, it was implemented and tested on breast cancer survivability and diabetics diagnostic modelling tasks respectively.

V. **Models and Tools Effectiveness -** To investigate the performance effectiveness of the proposed ensemble models and non-invasive analytical tools discussed in this thesis their performance outputs were compared with established benchmark models.

It has been argued that every ensemble model and learning method have their own benefits and limitations (Cordis., 2018; EPRS., 2019). Therefore, in the sub-sections that follow, discussion of the empirical investigations, the proposed models and learning methods, and the non-invasive predictive models are presented; in order to express their benefits, limitations and drawbacks.

6.2 Overview of the Models

This study incorporated factors from other well established theories and models applied in artificial intelligence (AI) and ensemble research streams. With this background, the purpose of this study include among others: improving the predictive performance of ensemble based models; testing and validating the performance of the proposed models. Applying and exploring the performance of the non-invasive predictive analytical tools that were developed based on the proposed and tested models in this thesis.

6.2.1 Benefits and Limitation of Ensemble Models

As demonstrated in this thesis, some of the motives why ensemble models are developed among others are:

I. After the models have been trained it can be applied on a wide range of tasks where the task conditions are similar or correlated. In addition, it is also possible to apply some predictive pre-conditions to the models.

II. Integrating sequential or EKF estimation techniques with ensemble models to address problem of lack of adequate historical data that is required to train the conventional ensemble models. Therefore, sequential learning methods or EKF estimation approaches are more reliable in the absence of large or missing values from the available data. It also performs better than batch (group of historical data) based learning techniques.

III. The possibility of integrating and enhancing the model once it has been trained and developed with visual interface models that can be used as a non-invasive analytic predictive models as demonstrated in this thesis.

6.2.2 Non Estimation Theory Based Models

When developing and testing algorithms availability of adequate historical data is very important in order to obtain good predictive performance. The predictive models described in Phase One, Phase Three, Phase Four and Phase Five were based on batch techniques (Rodríguez-Jiménez, et al., 2019). However, the EKF-RBFN in phase 2 was based on the concept of estimation theory. A non-estimation based learning models relies on available or historical data that can be divided into total dataset size at the time of training.

However, apart from the historical data requirements in batch learning methods there is a need for insights into the diversity of the individual models, decision making strategy and predictive committee members used in creating the models which the final ensemble models are based on.

The benefits of ensemble models that are based on non-estimation (batch-based) concept among other are:

I. The deep understanding of the data enables the model to create lower bias and variance. Therefore, it reduces classification errors by adjusting to the error rates of the weak hypotheses that forms the final predictive outcome.
II. The models have the capacity to reduce the spread or distribution of prediction therefore the overall model performance is better than single contribution model.
III. The models are capable of learning and remembering the input and output patterns of the dataset. Therefore, it can generate reasonable predictive output and applying trained models on unseen datasets. However, such unseen tasks must be similar or belong to the same domain as the original dataset used in training the model.
IV. The models are capable of representing the non-linear and complex topologies (in terms of time and space) that relates the input datasets to the predictive outcome results.

The limitations of ensemble models that are not based on non-estimation theory among others are:

I. Adequate and sufficient historical data are required to obtain a robust and reliable results.

II. The topology and computational complexity of the model coupled with the required training time and CPU is a major setback. As demonstrated in this research, the topology and computational complexity of standalone and ensemble models does not necessarily improve the predictive performance of the predictive models.

III. The model is static and cannot be used to simulate systems that are dynamic in nature or changes frequently in practice. The performance of the model is dependent on the undelaying loss function used in training and simulating the data and the weak classifiers.

IV. It could perform poorly when it is used on new tasks it has not learned from therefore, it may produce unexpected predictive outcomes that are not reliable and could not be used in non-invasive analytic predictive models.

V. The model is like a black box, therefore it cannot remember all the patterns and procedures in the training data. In addition the model is also sensitive to outliers. It is not tolerant to noise and can lead to under fitting and overfitting problems. It is demonstrated in this thesis by applying early stopping concept to lessen the limitation.

6.2.3 Estimation Theory Based Models

The predictive models described in Phase 2 were based on estimation theory. As demonstrated in the phase, it uses the Kalman filter algorithm to sequentially update and optimize parameters of RBFN models in order to enhance the performance of the integrated ensemble models. Firstly, one of the main reason of developing models that are sequential

(estimation theory) based is the unavailability of sufficient data to train the model. Secondly, the training data can only be collected and measured serially as it arrives. Thirdly, some data may be missing or arrived late than expected. Therefore, based on the previous and current states of the model, and environmental observations the previously predicted state can be corrected, the next state of the system can also be predicted and adjusted as new data arrive or are made available.

Some of the benefits of the models that were based on estimation theory among other are:

I. It can be used to model and simulate complex systems with little or no knowledge of the inner structure and complexity of the system.
II. After few iterations it is capable to adjust itself to the input and output structure of the system. It does this by using previous, current data and environment factors to correct and adjust itself.
III. Unlike the batch training method that is a case were the require data are available, it does not require large historical data to operate.
IV. It can be used to unravel complex and non-linear problems; as it is capable of linearizing the connection between the input and the output states of the system.

The drawbacks of some of the models that are based on estimation theory (UmaMageswari, et al., 2012; Lacambre, et al., 2013) among others include:

I. The linearization can lead to approximation errors which are not taken into consideration in the prediction and update steps. Therefore, it can underestimate the state uncertainties of the system.

II. The measurements that are applied in updating the weights and parameters of the system must be carried out at every iteration step of the system as new data arrives or are made available to the system. This can be time consuming and costly to achieve.

III. There is a possibility of generating unreliable output results when linearize on extreme volatile and irregular systems.

IV. The linear approximation of a nonlinear problem are based on assumptions which are not precise at all times as the environmental conditions changes.

6.3 Estimation Theory Based Versus Non Estimation Theory Based Models

The experimental investigations on ensemble models and learning methods, show that sequential learning technique should be applied when there is no adequate historical data. In a situation when the data only available consecutively, and the system to be modelled is complex and lacks insight. The investigations further demonstrates that batch learning methods should be applied when sufficient historical data to train the models are available, and there is sufficient knowledge of the process that can be used to model the system.

6.4 Chapter Summary

This chapter discussed and reflected on the proposed predictive models, and learning methods applied in this study. Firstly, the discussion highlights some of the motives why batch and sequential models are developed. It further discussed the benefits and drawbacks of the models' learning methods. Secondly, this chapter also presented key factors to be considered when selecting models, and training methods under different conditions and circumstances. The next chapter, therefore, moves on to discuss the conclusion future work.

Chapter 7: Conclusions and Future Work

7.1 Introduction

This chapter provides a conclusion to the results and discussions of the research presented in this thesis. The main motivation of the research has been to provide answers to the research hypothesis, develop new ensemble models with high prediction accuracy, and to provide an integrated non-invasive analytical tool based on the predictive models proposed in this thesis.

Following previous studies in ensemble modelling, this research has carried out investigations on how to develop different types of ensemble models with high prediction accuracy using different methods, conditions and tasks. The study investigated several aspects of AdaBoost methods, prognostic models, and predictive concepts to establish the features that affects the performance of standalone and ensemble methods.

7.2 Research Novelties, Contributions and Achievements

In this thesis extensive literature review and numerous empirical investigations were carried out. This was followed by the proposal, testing and validation of new ensemble predictive models, and non-invasive visual analytical predictive tools. Therefore, this session presents the research objectives that were accomplished during the study. It also highlights the research's novelty and contribution to knowledge.

I. **Objective One** - Review and empirical investigations in comparing performance of boosting methods, predictive performance of ensemble and standalone predictive techniques have been carried out. This has provided answers to Hypothesis 1 and Hypothesis 2 of this thesis.

II. **Objective Two** - An ensemble EKF-RBFN-AdaBoost that integrates EKF in training Radial Basis Function Networks (RBFN) in order to optimize the network parameters and applying AdaBoost as a meta-model to enhance stronger predictions accuracy has been proposed, developed and tested.

III. **Objective Three** - An ensemble model that is based on the concept of early stopping, and statistics data of the training samples: the mean, standard deviation and thresholds in order to minimise generalization error and to avoid overtraining that could result in overfitting has been proposed, developed and tested.

IV. **Objective Four** - An ensemble model that is based on the concepts of ensemble multivariate logistic regression that best predicts the binary response variable Y for the values of multiple X variables of the predictors have been proposed, developed and tested.

V. **Objective Five** - Analytical tools that were based on the theoretical concepts and practical application of the proposed ensemble models were used in proposing and developing non-invasive breast cancer survivability and diabetic diagnostic predictive analytical tools. The non-invasive analytical tools have been developed and tested.

VI. **Objective Six** - The performances of the ensemble models and the predictive analytical tools proposed and developed in this thesis have been tested and validated using benchmark samples.

VII. **Objective Seven** - Recommendations on the best approach for future work have been suggested. It has also been demonstrated that boosting method by

sampling performs relatively better than boosting by weighing. The thesis has therefore addressed this controversial issue in the literature.

7.3 Future Research and Recommendations

More broadly, this research has investigated different ensemble modelling methods, carried out experimental case studies, proposed, developed and tested new ensemble methods, multivariate logistic models and non-invasive analytical tools. However, it is recommended that further research can be undertaken in the following areas:

I. To advance the capabilities of the proposed EKF-RBFN-AdaBoost model by extending the model so that it can handle multiple base classifiers and datasets for users to select from. This could involve the development of a model reference control interface such as GUI toolkits. This would enable users to load different datasets, select the required base learners, and carry out pre-training algorithmic settings before training.

II. It would also be interesting to investigate further the effects of diversity of weak classifiers on performance of the proposed models.

III. To explore and investigate the potential and predictive effects of imbalanced and complex training tasks on the proposed models.

IV. To investigate the application of UKF (Unscented Kalman Filter), UPF (Unscented Particle Filter) in training the RBFN as these methods could perform better than EKF in terms of estimation accuracy (György, et al., 2014; Mariani & Ghisi, 2007).

V. To further investigate algorithmic factors that could affect performance of ensemble classification accuracy.

7.4 Conclusion

One of the main aims and objectives of this research is to improve the prediction accuracy of ensemble models. Following reviews of previous work on ensemble modelling, this research has explored some of the existing gaps in the literature and how to fill them.

Based on current issues of ensemble models: empirical investigations were carried out to provide answers to the research hypothesis. Then three different predictive modes, and two predictive analytical tool were proposed and developed.

The proposed models and tools models were tested on Breast Cancer survivability, Diabetes diagnosis, and Staff Absenteeism, Credit Card Default Payment and many other datasets. The results of the tests on the proposed models shows improved prediction accuracy of the models when compared with a number of benchmarked standalone and ensemble models.

Some of the findings and conclusions of this thesis are:

I. The study has shown and confirmed that the complexity, the topology of algorithm, and time required to train the tasks does not necessarily improve the performance and accuracy of the models. Even though more difficult models such as Neural Network and Support Vector Machine requires more time, tuning efforts, and CPU resources while training.

II. The results of this study indicate that implementing AdaBoost by resampling method perform slightly better than implementing the model by reweighting method. It was found that the average performance of correctly classified by reweighting method is 73% compared with boosting by resampling method which was 74%.

III. The study has gone some way towards enhancing the understanding and the possibility of integrating EKF, RBFN and AdaBoost as an optimized predictive model.

IV. The results of the studies further demonstrate that the prediction accuracies of some of the proposed models are more accurate compared to other benchmark models used during the research.

V. This study has found that the proposed EKF-RBFN-AdaBoost, and the early stopping concepts models predict quite faster after few iterations. It also converges faster compare to other conventional algorithms used in this study.

VI. This study has demonstrated the usefulness of the ensemble models proposed in this thesis by proposing, developing and testing non-invasive analytical tools. The tools links the gaps between theoretical concepts and practical application of the proposed models to predict breast cancer survivability and diabetic diagnostics in patients. The tool is privacy protective and non-invasive, cost effective and could be used over the internet.

VII. Some of the findings of this study have been peer reviewed, presented and published in five different international proceedings, journals and chapters in books.

Appendix A – Time Line Comparison and Summary of Some of AdaBoost variants

Table A.1 Comparison Summary of Common AdaBoost Variants and Related Boosting Algorithms

Variants	**Modified Variant**	**Application**	**Binary/ Multi-class**	**Base Classifier**	**Benefits and Advantages**	**Limitations and Problems**	**Loss Function**	**Remarks / Comments**
AdaBoost (Freund & Schapire, 1995; Freund & Schapire, 1997)	-	Image recognition	Yes	Decision Stump	Reduces classification error	The performance depends on the on the weak learners generated during boosting process.	Exponential function	Generates classifiers whose performance is a little better than random guess
AdaBoost.M1 (Freund & Schapire, 1996)	AdaBoost	Image /Photography and number recording	Yes /Yes	Decision Stump	It adjusts to the error rates of the weak hypotheses	Performs poorly on noisy datasets.	Exponential function	The weak learner is measured only in terms of error rate
AdaBoost.M2 (Freund & Schapire, 1996)	AdaBoost.M1	Digital and Electronic recognition	Yes / Yes	C4.5	Corrects AdaBoost.M1 limitation on maximum error weight.	Overfitting/Slow convergence rate	Exponential function	It has same performance in minimizing the training and test error rates
LogitBoost (Friedman, et al., 2000), (Friedman, 1997)	AdaBoost.M1	Credit Appraising/Traceability	Yes /Yes	Decision Stump	Handle noisy data/Perform better than AdaBoost.M1/Fast Computation	It produces complex regression tree	Logistic loss	Uses logistic regression techniques /Newton algorithm to fit additive logistic regression model
Real AdaBoost (Schapire & Singer, 1999) (Friedman, et al., 2000)	AdaBoost	Face Detection	Yes/No	Decision Stump	Weighted probability / additive logistic model	Numerical and overfitting problems.	Additive logistic regression.	The weak learner returns a class probability. The contribution to the final classifier is half the logit transform of the probability.

BrownBoost (Freund, 2001)	AdaBoost	Face Detection	Yes /No	Non-convex loss function	Handle multi-class tasks	Convergence problem/Noisy resistant	Differential equations/ Brownian motion	It is an adaptive version of BBM[8]
FloatBoost (Li, et al., 2003) (Li & Zhang, 2004)	Floating search /AdaBoost	Face detection systems /Face detection	Yes /No	Decision Stump	Error rate / fewer weaker classifiers	Overhead costs/Deletes unfavourable weak classifiers	Incorporates Floating Search	Uses input-dependent regularisers
WeightBoost (Jin, et al., 2003)	AdaBoost	Text categorization	Yes/No	Decision Stump	Overfitting / noisy data problem	More robust than AdaBoost on noisy data and performs better	Exponential function	It guaranteed to minimize training errors.
Modest AdaBoost (Vezhnevets & Vezhnevets, 2005)	Real AdaBoost and Gentle AdaBoost	Object detection	Yes/No	Decision Stump	Less generalization error compares with Gentle AdaBoost and Real AdaBoost	Unstable performance	It uses inverted distributions scheme	It has a higher training error
Ent-Boost (Le & and Satoh, 2007)	AdaBoost	Computer vision / face detector	Yes/No		It has good performance and compact storage space	Expensive and cumbersome stopping criterion	Entropy function	It uses class entropy information to estimate optimal number of bins
ReweightBoost (Rodríguez & Maudes, 2008)	AdaBoost	Computer vision / face detector	Yes /No	ADA[9]	Computational complexity	Uses few base classifiers /Rely on parameters		It decreases the generalization error
Gentle AdaBoost (Wu & Nagahashi, 2014)	Real AdaBoost	Object detection	Yes /No	Decision Stump	Reliable/stable than Real AdaBoost/ Outperforms Real AdaBoost and LogitBoost	Overfitting and high computational cost	Newton stepping /Least-square regression	It uses Newton stepping instead of exact optimization at each step

[8] Boosting-By-Majority

[9] Adaptive Discriminant Analysis

References

Abuhasel, K. A., Iliyasu, A. M. & Fatichah, C., 2015. A combined AdaBoost and NEWFM Technique for Medical Data classification. *Lecture Notes in Electrical Engineering,* 18 February, Volume 339, pp. 801-809.

Abuhasel, K., Iliyasu, A. & Fatichah, C., 2015. A Combined AdaBoost and NEWFM Technique for Medical Data Classification. *Information Science and Applications,* Volume 339, pp. 801-809.

Adegoke, V., Banissi, E. & Barikzai, S., 2019. Improving Prediction Accuracy of Breast Cancer Survivability and Diabetes Diagnosis via RBF Networks trained with EKF models. *International Journal of Computer Information Systems and Industrial Management Application,* 11(2019), pp. 082-100.

Adegoke, V., Chen, D. & Banissi, E., 2017. *Prediction of breast cancer survivability using ensemble algorithms.* Osijek, Croatia, IEEE.

Adegoke, V., Chen, D., Banissi, E. & Barikzai, S., 2017. *Prediction of breast cancer survivability using ensemble algorithms.* Osijek, Croatia, IEEE.

Adegoke, V., Chen, D., Banissi, E. & Barikzai, S., 2019. *Enhancing Ensemble Prediction Accuracy of Breast Cancer Survivability and Diabetes Diagnostic using optimized EKF-RBFN trained prototypes.* Porto, Portugal, Springer.

Adegoke, V., Chen, D., Barikzai, S. & Banissi, E., 2017. *Predictive Ensemble Modelling: Experimental Comparison of Boosting Implementation Methods.* s.l., s.n.

Agarwal, M., 2020. *Inc42.* [Online]

Available at: https://inc42.com/resources/understanding-the-organisational-culture-iceberg-to-drive-success/

[Accessed 12 December 2021].

Alghamdi, M. et al., 2017. Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: The Henry Ford ExercIse Testing (FIT) project. *PLoS One,* 12(7).

Alharthi, H., 2018. Healthcare predictive analytics: An overview with a focus on Saudi Arabia. *Journal of Infection and Public Health,* 11(6), pp. 749-756.

Allen, M., 1997. The problem of multicollinearity. In: *Understanding Regression Analysis.* s.l.:s.n.

Alpaydin, E., 2010. *Introduction to Machine Learning : Adaptive Computation and Machine Learning.* 2nd ed. s.l.:MIT Press, Cambride, M.A.

Alpenberg, J. & Scarbrough, D. P., 2016. Exploring communication practices in lean production. *Journal of Business Research,* 69(11), p. 4959–4963.

Aly, H., 2020. An intelligent hybrid model of neuro Wavelet, time series and Recurrent Kalman Filter for wind speed forecasting. *Sustainable Energy Technologies and Assessments.*

Ambati, V., 2021. *Everything about Ensemble Learning.* [Online]

Available at: https://medium.com/geekculture/everything-about-ensemble-learning-46cd96d5498e

[Accessed 03 August 2021].

American-Cancer-Society, 2015. *Breast Cancer Facts & Figures 2015-2016.* [Online] Available at: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2015-2016.pdf [Accessed 05 May 2017].

Anjum., S., 2013. Algorithms for Predictive Classification in Data Mining - A comparison of Evaluation methodologies. *Journal of Industrial and Intelligent Information,* 1(2).

Anning-Dorson, T., Odoom, R., Acheampong, G. & weneboah-Koduah, E., 2017. Innovation and organizational development: the role of organizational leadership. *African journal of economic and management studies,* 8(3), pp. 338-351.

Anon., 2002. Sarimveis. H.;Alexandridis, A.;Mazarakis, S.;Bafas, G.;. *A New Algorithm for developing Dynamic Radial Basis Function Neural Network Models based on Genetic Algorithms,* Volume 10, pp. 949-954.

Anon., 2005. *Data mining : Practical Machine Learning Tools and Techniques.* 2nd ed. s.l.:The Morgan Kaufmann Series in Data mining Systems..

Anon., 2019. Evaluating Overfit and Underfit in Models of Network Community Structure. *IEEE Trans. Knowledge and Data Engineering,* 32(9), pp. 1722-1735.

Anon., 2020. Harmless Overfitting: Using Denoising Autoencoders in Estimation of Distribution Algorithms. *Journal of Machine Learning Research,* Volume 21, pp. 1-31.

Apreutesei, N. A. et al., 2018. Predictions of ocular changes caused by diabetes in glaucoma patients. *Computer Methods and Programs in Biomedicine,* Volume 154, pp. 183-190.

Arora, S. & Hazan, .. E., 2016. *Machine learning and Artifical Intelligence.* [Online] Available at: https://www.cs.princeton.edu/courses/archive/fall16/cos402/lectures/402-lec5.pdf
[Accessed 30 August 2021].

Arsov, N., Pavlovski, M., Basnarkov, L. & Kocarev, L., 2017. Generating highly accurate prediction hypotheses through collaborative ensemble learning. *Scientific Reports .*

Attarian, A., Batze, J., B.Matzuka & Tran, H., 2012. Application of the Unscented Kalman Filtering to Parameter Estimation. *Mathematical Modeling and Validation in Physiology*, 11 September, Volume 2064, pp. 75-88.

Awad, M. & Khanna, R., 2015. Support Vector Machines for Classification. In: *Efficient Learning Machines.* Berkeley, CA: Apress.

Bagga, G., 2013. How to keep the talent you have got?. *Human Resource Management International Digest,* 21(1), pp. 3-4.

Baharin, N. & Hanafi, W., 2018. Effects of Talent Management on Employee Retention: A Case Study of Hospitality Industry. *Global business and management research,* 10(2).

Barakat, N., Bradley, A. P. & Barakat, M. N. H., 2010. Intelligible Support Vector Machines for Diagnosis of Diabetes Mellitus. *Transactions on Information Technology in Biomedicine,* 14(4).

Batuwita, R. & Palade, V., 2013. Class imbalance learning methods for support vector machines. In: H. Haibo & M. YunQian, eds. *Imbalanced Learning: Foundations, Algorithms, and Applications.* Berlin, Germany,: John Wiley & Sons, pp. 83-99.

BBC, 2014. *Paramedics leaving London Ambulance service trebles.* [Online]
Available at: https://www.bbc.co.uk/news/uk-england-london-28158180
[Accessed 12 December 2021].

BBC, U., 2018. *Labour's Tom watson 'reversed' type-2 diabetes through diet and exercise,* London: British Broadcasting Corporation.

Berry, M. & Linoff, G., 2011. In: *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management.* 2nd ed. s.l.:Wiley Publishing Inc.

BHF, 2018. *CVD Statistics – BHF UK Factsheet,* s.l.: BHF(British Heart Foundation).

Bian, S., 2007. On diversity and accuracy of homogeneous and heterogeneous ensembles. *International journal of hybrid intelligent systems,* 4(2), pp. 103-128.

Bishop, C., 2006. *Pattern Recognition and Machine Learning.* New York, NY. USA: Springer.

Blagus, R. & Lusa, L., 2013. SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics,* 14(106).

Bohler, L., Ritzberger, D., Hametner, C. & Jakubek, S., 2021. Constrained extended Kalman filter design and application for on-line state estimation of high-order polymer electrolyte membrane fuel cell systems. *International Journal of Hydrogen Energy,* 46(35), pp. 18604-18614.

Bonab, H. & Can, F., 2016. *A Theoretical Framework on the Ideal Number of Classifiers for Online Ensembles in Data Streams.* s.l., s.n., pp. 2053-2056.

Bonab, H. & Can, F., 2018. Less Is More: A Comprehensive Framework for the Number of Components of Ensemble Classifiers. *Transactions on Neural Networks and learning Systems,* 14(8).

Botvinick, M. et al., 2019. Reinforcement Learning, Fast and Slow. *Trends in Cognitive Sciences,* 23(5).

Bradley, A., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition,* 30(7), pp. 1145-1159.

Braun, S., Peus, C., Weisweiler, S. & Frey, D., 2013. Transformational leadership, job satisfaction, and team performance: A multilevel mediation model of trust. *The Leadership quarterly,* 24(1), pp. 270-283.

Breiman, L., 1996. Bagging predictors. *Machine Learning,* 24(2), pp. 123-140.

Brown, G., Wyatt, J. & Tino, P., 2005. Managing Diversity in Regression Ensembles. *Journal of Machine Learning Research,* pp. 1621-1650.

Butler, K., Fried, M. & Bauer, K., 2018. The effectiveness of using diversity to select multiple classifier systems with varying classification thresholds. *Journal of Algorithms & Computational Technology.*

Cao, J., Kwong, S. & Ran, w., 2012. A noise-detection based AdaBoost algorithm for mislabeled data. *Pattern Recognition,* 45(12), pp. 4451-4465.

Chandra, A., Chen, H. & Yao, X., 2006. Trade-Off Between Diversity and Accuracy in Ensemble Generation. In: Y. Jin, ed. *Multi-Objective Machine Learning. Studies in Computational Intelligence.* s.l.:Springer, Berlin, Heidelberg.

Chang, C., Li, C. & Wang, H., 2007. *Using Principal Component Analysis and Radial Basis Function Neural Networks for Image Contents Classification,.* s.l., Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 7-10.

Chen, D., Guo, K. & Li, B., 2019. Predicting Customer Profitability Dynamically over Time: An Experimental Comparative Study. In: H. Hernández & N. Milián, eds. *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. CIARP 2019. Lecture Notes in Computer Science.* s.l.:Springer, Cham.

Chen, D. et al., 2021. *Deep Learning Causal Attributions of Breast Cancer.* London, Springer, Cham.

Chen, D. et al., 2021. Deep Learning Causal Attributions of Breast Cancer. In: A. Kohei, ed. *Intelligent Computing. Lecture Notes in Networks and Systems.* s.l.:Springer, Cham.

Cheng, H. & Tan, P.-H., 2010. Efficient Algorithm for Localized Support Vector Machine. *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING,* 22(10).

Cheng, H., Tan, P.-H. & Jin, R., 2007. *Localized Support Vector Machine and Its Efficient Algorithm.* s.l., s.n., pp. 461-466.

Chernodub, A., 2014. Training Neural Networks for classification using the Extended Kalman Filter: A comparative study. *Optical Memory and Neural Networks,* 23(2), p. 96–103.

Chernodub, A., 2014. Training Neural Networks for classification using the Extented Kalman Filter: A Comparative Study. *Optimal and Neural Networks (Information Optics),* 23(2), pp. 96-103.

Chernodub, A. N., 2014. Training Neural Networks for classification using the Extended Kalman Filter: A comparative study. *Optical Memory and Neural Networks,* 23(2), pp. 96-103.

Chou, J.-S., Cheng, M.-Y. & Wu, Y.-W., 2013. Improving classification accuracy of project dispute resolution using hybrid artificial intelligence and support vector machine models. *Expert Systems with Applications,* 40(6), pp. 2263-2274.

Cogburn, R., 2019. *Fobes.* [Online]
Available at: https://www.forbes.com/sites/forbestechcouncil/2019/05/13/the-past-present-and-future-of-automated-predictive-technology/#b43a4455317e

Colozza, M. et al., 2005. Proliferative markers as prognostic and predictive tools in early breast cancer: where are we now?. *Ann Oncol.,* 16(11), pp. 1723-39.

Comaford, C., 2018. *Three Steps To Fix A Broken Culture.* [Online]
Available at: https://www.forbes.com/sites/christinecomaford/2018/05/12/three-steps-to-fix-a-broken-culture/?sh=2ac44f9e42f0
[Accessed 12 December 2021].

Cordis., 2018. *Understanding the powers and limitations of algorithms to solve logic formula.* [Online]
Available at: https://cordis.europa.eu/article/id/250876-understanding-the-powers-and-limitations-of-algorithms-to-solve-logic-formulas
[Accessed 13 August 2021].

Corinna, C. & Vladimir, V., 1995. Support-vector networks. *Machine Learning,* 20(3), pp. 273-297.

Cover, T. & Thomas, J., 2006. *Elements of Information Theory.* 2nd ed. s.l.:Wiley.

Dark, F., Whiteford2, H. & Ashkanasy, N., 2017. The impact of organisational change and fiscal restraint on organisational culture. *International Journal of Mental Health Systems,* 11(11).

Day, O. & Khoshgofaar, T., 2017. A survey on heterogeneous transfer learning. *Journal of Big Data,* 4(29).

Delen, D., Walker, G. & Kadam, A., 2005. Predicting breast cancer survivability: A comparison of three data mining methods. *Artif Intell Med.,* 34(2), pp. 113-27.

Demiriz, A., Bennett, K. & Shawe-Taylor, J., 2002. Linear programming boosting via column generation. *Machine Learning,* 46(1-3), pp. 225-254.

Dietterich, T., 2000. *Ensemble methods in machine learning.* s.l., s.n., pp. 1-15.

Dróżdż, K. & Szabat, K., 2016. *Application of Unscented Kalman Filter in adaptive control structure of two-mass system.* s.l., IEEE.

Dua, D. & Karra Taniskidou, E., 2017. *UCI Machine Learning Repository,* Irvine, CA: Center for Machine Learning and Intelligent Systems.

Duda, R. O., Hart, P. E. & Stork, D. G., 2000. *Pattern classification.* 2nd ed. s.l.:John Wiley and Sons, INC.

Dunlavy, D. & Gilpin, S., 2018. *Heterogeneous Ensemble Classifiers,* s.l.: s.n.

Duvenaud, D., Maclaurin, D. & Adams, R., 2016. *Early Stopping as Nonparametric Variational Inference.* s.l., s.n., pp. 1070-1077.

Eisenbeiß, S. & Boerner, S., 2013. A Double-edged Sword: TransformationalLeadership and Individual Creativity. *British Journal of Management,* Volume 24, pp. 54-68.

EPRS., 2019. *Understanding algorithmic decision-making: Opportunities and challenges.* [Online]
Available at:
europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU(2019)624261_EN.pdf
[Accessed 12 August 2020].

Fan, M., Zheng, B. & Li, L., 2015. A novel Multi-Agent Ada-Boost algorithm for predicting protein structural class with the information of protein secondary structure. *J Bioinform Comput Biol,* 13(5).

Feng, W., Huang, W. & Ren, J., 2018. Class Imbalance Ensemble Learning Based on the Margin Theory. *Appl. Sci.,* 8(815).

Fitzmaurice, C., Dicker, D. & al, A. P. e., 2015. The Global Burden of Cancer 2013. *JAMA Oncology,* Volume 1(5):690, pp. 505-527.

Formenti, S., Arslan, A. & Love, S., 2012. Global Breast Cancer: The Lessons to Bring Home. *International Journal of Breast Cancer.*

Freund, Y., 2001. An Adaptive Version of the Boost by Majority Algorithm. *Machine Learning,* Volume 43, p. 293–318.

Freund, Y. & Schapire, R., 1977. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of computer and system sciences,* Volume 55, pp. 119-139.

Freund, Y. & Schapire, R., 1996. *Experiments with a New Boosting Algorithm.* Bari, Italy, Morgan Kaufmann Publishers Inc. San Francisco, pp. 148-156.

Freund, Y. & Schapire, R., 1996. *Experiments with a New Boosting Algorithm.* s.l., Morgan Kaufmann, pp. 148-156.

Freund, Y. & Schapire, R., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences 55,* pp. 119-139.

Freund, Y. & Schapire, R., 1997. A decision-theoretic generalization of on-line learning and application to boosting. *Journal of computer and system sciences,* Volume 55, pp. 119-139.

Freund, Y. & Schapire, R., 1999. A Short Introduction to Boosting. *Journal of Japanese Society for Artificial Intelligence,* 4(5), pp. 771-780.

Freund, Y. & Schapire, R. E., 1995. *A decision-theoretic generalization of on-line learning.* s.l., Springer-Verlag, p. 23–37.

Freund, Y. & Schapire, R. E., 2014. *Boosting Foundations and Algorithms.* s.l.:MIT Press.

Friedman, J., 1997. Greedy function approximation : A gradient boosting machine. *Journal of Computer and System Sciences,* 55(1), pp. 119-139.

Friedman, J., Hastie, T. & Tibshirani, R., 2000. Additive logistic regression: a statistical view of boosting. 28(2), pp. 337-374.

Friedman, J., Hastie, T. & Tibshirani, R., 2000. Special Invited Paper Additive Logistic Regression: A Statistical View of Boosting. *The Annals of Statistics,* 28(2), pp. 337-407.

Friedman, J., Hasti, T. & Tibshirani, R., 2000. Special Invited Paper. Additive Logistic Regression: A Statistical View of Boosting. *The Annals of Statistics,* 28(2), pp. 337-374.

Galar, D. & Kumar, U., 2017. Chapter 5 - Diagnosis. In: D. Galar & U. Kumar, eds. *eMaintenance Essential Electronic Tools for Efficiency.* s.l.:Academic Press, pp. 235-310.

Galar, M., Fernandez, A., E.Barrenechea & H.Bustince, 2012. A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS - PART C: APPLICATIONS AND REVIEWS,,* Volume 4.

Galar, M., Fernandez, A. & et.al, 2012. A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches. *Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) ,* 42(4).

Gao, Q. & Xu, X., 2014. *The analysis and research on computational complexity.* s.l., s.n., pp. 3467-3472.

Gao, W. & Zhou, Z.-H., 2013. Onthedoubtaboutmarginexplanationofboosting. *Artificial inytelligence,* Volume 203, pp. 1-18.

Géron, A., 2018. *Neural networks and deep learning.* s.l.:O'Reilly Media, Inc.

Geyfman, V., 2014. Engaging Government Employees: Motivate and Inspire Your People to Achieve Superior Performance. *The Journal of applied management and entrepreneurshi,* pp. 121-123.

Ghassemi, M. et al., 2020. A Review of Challenges and Opportunities in Machine Learning for Health. *AMIA Jt Summits Transl Sci Proc,* p. 191–200.

Ghojogh, B. & Crowley, M., 2019. *University of Waterloo - The theory behind overfitting, cross validation, regularization, bagging, and boosting: tutorial.* [Online]
Available at: https://uwaterloo.ca/scholar/mcrowley/publications/theory-behind-overfitting-cross-validation-regularization-bagging-and-boosting
[Accessed 26 July 2021].

Gladwin, L., 1997. Alan Turning, Enigma and the breaking of German Machine Ciphers in World War II. *Turning and Enigma,* pp. 202-217.

Glaze, B., 2019. *Tens of thousands of ambulance staff have quit the NHS in last nine years.* [Online]
Available at: https://www.mirror.co.uk/news/politics/tens-thousands-ambulance-staff-quit-15005720
[Accessed 8 December 2021].

GT, 2020. *Biobot Analytics Pulls $4.2M for Sewage Testing Technology.* [Online]
Available at: https://www.govtech.com/biz/Biobot-Analytics-Pulls-42M-for-Sewage-Testing-Technology.html

Guerci, J., Goetz, R. & J., 1994. A method for improving extended Kalman filter performance for angle-only passive ranging. *Transactions on Aerospace and Electronic Systems,* October.pp. 1090-1093.

Gunawardana, A. & Shani, G., 2009. A Survey of Accuracy Evaluation Metrics of Recommendation Tasks. *Journal of Machine Learning Research,* Volume 10, pp. 2935-2962.

Guo, H., Zhou, J. & Wu, C.-a., 2020. Ensemble learning via constraint projection and undersampling technique for class-imbalance problem. *Soft Computing,* Volume 24, p. 4711–4727.

Guo, J. et al., 2013. Machine Learning Predictions for Underestimation of Job Runtime on HPC System. In: R. Yokota & W. Wu, eds. *Supercomputing Frontiers - 4th Asian Conference, SCFA 2018* . Singapore: Springer, pp. 187-206.

Guradian, 2015. *London Ambulance has 'embedded culture of bullying and harassment', independent report finds.* [Online]
Available at: https://www.yourlocalguardian.co.uk/news/13368659.london-ambulance-has-embedded-culture-of-bullying-and-harassment-independent-report-finds/
[Accessed 13 December 2021].

György, K., Kelemen, A. & Dávid, ,., 2014. Unscented Kalman Filters and Particle Filter Methods for Nonlinear State Estimation. *Procedia Technology,* Volume 12, pp. 65-74.

Hajiabadi, H., Molla-Aliod, D., Monsefi, R. & Yazdi, S., 2020. Combination of loss functions for deep text classification. *International Journal of Machine Learning and Cybernetics,* 11(4), pp. 751-761.

Hamze-Ziabaria, S. & Bakhshpoorib, T., 2017. Improving the prediction of ground motion parameters based on an efficient bagging ensemble model of M5' and CART algorithms. *Applied Soft Computing.*

Hastie, T., Tibshirani, R. & Friedman, J., 2009. *The Elements of Statistical Learning: Data mining, Interface and Predictions Springer.* 2nd ed. s.l.:Springer.

Haykin, S., 1996. *Adaptive Filter Theory.* 3rd ed. s.l.:Prentice-Hall,.

Haykin, S., 2008. *Neural Networks and Learning Machines: A Comprehensive Foundation.* Third ed. s.l.:Prentice Hall.

Hbs, 2020. *5 TIPS FOR MANAGING CHANGE IN THE WORKPLACE.* [Online]
Available at: https://online.hbs.edu/blog/post/managing-change-in-the-workplace
[Accessed 10 December 2021].

He, H., McGinnity, T., Coleman, S. & Gardiner, B., 2014. Linguistic Decision Making for Robot Route Learning. *IEEE Transactions on Neural Networks and Learning Systems*, January, 25(1), pp. 203-215.

HernáNdez-Lobato, D., MartíNez-MuñOz, G. & SuáRez, A., 2013. How large should ensembles of classifiers be?. *Pattern Recognition,* 46(5), pp. 1323-1336.

He, Y. et al., 2018. Kalman Filtering Algorithm for Systems with Stochastic Nonlinearity Functions, Finite-Step Correlated Noises, and Missing Measurements. *Discrete Dynamics in Nature and Society,* Volume 2018.

Ho, T., 1998. The random space method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 20(8), pp. 832-844.

IBM, 2021. *Underfitting.* [Online]
Available at: https://www.ibm.com/cloud/learn/underfitting
[Accessed 02 August 2021].

ICL., 2021. *Artificial Neural Networks.* [Online]
Available at: https://www.doc.ic.ac.uk/~nuric/teaching/imperial-college-machine-learning-

neural-networks.html

[Accessed 2 August 2021].

Igareta, A., 2021. *The Million-Dollar Question: When to Stop Training your Deep Learning Model.* [Online]

Available at: https://towardsdatascience.com/the-million-dollar-question-when-to-stop-training-deep-learning-models-fa9b488ac04d

[Accessed 20 August 2021].

Imhoff, R. & Lamberty, P., 2020. A Bioweapon or a Hoax? The Link Between Distinct Conspiracy Beliefs About the Coronavirus Disease (COVID-19) Outbreak and Pandemic Behavior. *Social Psychological and Personality Science.*

Indeed., 2021. *What Is the Strategic Planning Process?.* [Online]

Available at: https://www.indeed.com/career-advice/career-development/strategic-planning-process

[Accessed 12 December 2021].

Ishak, A. & Tokhi, M., 2017. Modelling and Control of Standing Up and Sitting Down. *Pertanika J. Sci. & Technol,* 25(S), pp. 243-250.

ITV, 2018. *Surge in heart attacks and strokes predicted as diabetes epidemic takes its toll,* London, UK: ITV Report.

Jadhav, S. & He, H., 2015. *Prediction of Earnings per Share for Industry.* Lisbon, Portugal, IEEE, pp. 425-432.

Jain, A., Duin, R. & Mao, J., 2000. Statistical Pattern Recognition: A Review. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, January, 22(1).

James, G., Witten, D. & Hastie, T. R., 2014. *An Introduction to Statistical Learning with Applications in R.* New York : Springer.

Javid, T., Gupta, M. & Gupta, A., 2020. A hybrid-security model for privacy-enhanced distributed data mining. *Journal of King Saud University - Computer and Information Sciences.*

Jeon, Y.-S., Yang, D.-H. & Lim, D.-J., 2019. FlexBoost: A Flexible Boosting Algorithm With Adaptive Loss Functions. *IEEEAccess.*

Jerome, N., 2013. Empirical Investigation of the Impact of Organizational Culture on Human Resource Management. *International Journal of Business and Social Science,* 4(5).

Jimenez, D., 1998. *Complexity Analysis The university of Texas at Austin.* [Online] Available at: https://www.cs.utexas.edu/users/djimenez/utsa/cs1723/lecture2.html [Accessed 27 July 2021].

Jin, R. et al., 2003. *A new boosting algorithm using inputdependent.* s.l., s.n.

Johennesse, C. & Chou, T.-K., 2017. Employee Perceptions of Talent Management Effectiveness on Retention. *lobal business and management research,* 9(4), p. 46.

Joshi, R. & Dhakai, C., 2021. Predicting type 2 Diabetes using Logistic Regressio and machine Learning Approaches. *International Journal of environmental Resaerch and Public Health,* 18(14).

Kahane, A., 2017. Stretch collaboration: how to work with people you don't agree with or like or trust. *Strategy & Leadership,* 45(2), pp. 42-45.

Kalman, R., 1960. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME-Journal of Basic Engineering,* 82(D), pp. 35-45.

Kamath, H. et al., 2011. *RBF and BPNN Combi Model Based Filter Application for maximum Power Point Tracker of PV Cell.* Hong Kong, s.n.

Karimi, P. & Jazayeri-Rad, H., 2014. Comparing the fault diagnosis performances of single Neural Networks and two Ensemble Neural Networks based on the boosting methods. *Journal of Automation and Control,* 2(1), pp. 21-32.

Karimi, P. & Jazayeri-Rad, H., 2014. Comparing the Fault Diagnosis Performances of Single Neural Networks and Two Ensemble Neural Networks Based on the Boosting Methods. *Journal of Automation and Control,* 2(1), pp. 21-32.

Karim, R., 2018. *Intuitions on L1 and L2 Regularisation.* [Online]
Available at: https://towardsdatascience.com/intuitions-on-l1-and-l2-regularisation-235f2db4c261#15c2
[Accessed 02 August 2021].

Kazienko, P., Lughofer, E. & Trawiński, B., 2013. Hybrid and Ensemble Methods in Machine Learning. *Journal of Universal Computer Science,* 19(4), pp. 457-461.

Ke, G. & et.al., 2017. *LightGBM: A Highly Efficient Gradient Boosting.* s.l., NIPS'17, pp. 3149-3157.

Kelly, C. et al., 2019. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med,* p. 185.

Kelwade, J. & Salankar, S., 2016. *Prediction of heart abnormalities using Particle Swarm Optimization in Radial Basis Function Neural network.* s.l., IEEE, pp. 793 - 797.

Kim, K., Choi, H.-I. & Oh, K., 2017. Object detection using ensemble of linear classifiers with fuzzy adaptive boosting. *EURASIP Journal on Image and Video,* 2017(40).

Kotu, V. & Deshpande, B., 2019. Data Science. In: *Chapter 2 - Data Science Process.* s.l.:Morgan Kaufmann, pp. 19-37.

Krok, A., 2013. The Development of Kalman Filter Learning Technique for Artifical Neural Networks. *Journal of Telecommunications and Information Technolgy,* pp. 16-21.

Kumar, G., Banerjee, R. N., Meena, P. L. & Ganguly, K., 2016. Collaborative culture and relationship strength roles in collaborative relationships: a supply chain perspective. *Journal of Business & Industrial Marketing,* 31(5), pp. 587-599.

Kuncheva, L. I., 2014. *Combining Pattern Classifiers: Methods and Algorithm.* 1st ed. s.l.:John Wiley & Sons.

Kuncheva, L. I. & Whitaker, C. J., 2003. Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy. *Machine Learning,* Volume 51, p. 181–207.

Kuncheva, L. & Whiitaker, C., 2003. Measures of Diversity in Classifier Ensembles. *Machine Learning,* Volume 51, pp. 181-207.

Kuncheva, L. & Whitaker, C., 2003. Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy. *Machine Learning ,* 51(https://doi.org/10.1023/A:1022859003006), p. 181–207.

Kurban, T. & Beşdok, E., 2009. A Comparison of RBF Neural Network Training Algorithms for Inertial Sensor Based Terrain Classification. 12 August, Volume 9, pp. 6312-6329.

Kurbban, T. & Besdok, E., 2009. A Comparison of RBF Neural Network Training Algorithms for Inertial Sensor Based Terrain Classification. *Sensors (Basel),* 9(8), pp. 6312-29.

Kwon, S. & Lee, S., 2016. Recent Advances in Microwave Imaging for Breast Cancer Detection. *International Journal of Biomedical Imaging.*

Lacambre, J., Narozny, M. & Louge, J., 2013. *Limitations of the unscented Kalman filter for the attitude determination on an inertial navigation system.* s.l., IEEE, pp. 187-192.

Lacey, T., 2018. *Tutorial: The Kalman Filter.* [Online]
Available at:
http://web.mit.edu/kirtley/kirtley/binlustuff/literature/control/Kalman%20filter.pdf
[Accessed 2 March 2018].

Lacey, T., 2020. *Tutorial - The Kalman Filter.* [Online]
Available at:
http://web.mit.edu/kirtley/kirtley/binlustuff/literature/control/Kalman%20filter.pdf
[Accessed 04 February 2021].

Langer, A. M., 2016. *Guide to Software Development: Designing and Managing the Life Cycle.* 2nd ed. s.l.:Springer.

LAS., 2021. *About us.* [Online]

Available at: https://www.londonambulance.nhs.uk/about-us/

[Accessed December 20 2021].

LAS, 2021. *What we do.* [Online]

Available at: https://www.londonambulance.nhs.uk/about-us/what-we-do/

[Accessed 12 December 2021].

Latinne, P., Debeir, O. & Decaestecker, C., 2001. Limiting the number of trees in random forests. In: *Multiple Classifier Systems. MCS 2001. Lecture Notes in Computer Science.* s.l.:Springer, Berlin, Heidelberg.

Lauer, F. & Bloch, G., 2005. Ho–Kashyap classifier with early stopping for regularization. *Pattern Recognition Letters,* 27(9), pp. 1037-1044.

Le, D. & and Satoh, S., 2007. *Ent-Boost : Boosting Using Entropy Measure for Robust Object Detection.* s.l., s.n.

Lee, .., 2018. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Algorithms in Culture,* 5(1).

Lee, D., Jeng, S. & Chandrasekar, P., 2004. Applying data mining techniques for traffic incident analysis. *Journal of The Institution of Engineers,* 44(4).

Lee, H. & Cummings, G., 2008. Factors influencing job satisfaction of front line nurse managers: a systematic review. *Journal of nursing management,* 16(7), pp. 768-783.

Lee, S.-J., Chen, T., Yu, L. & Lai, C.-H., 2018. Image Classification Based on the Boost Convolutional Neural Network. *IEEE Open Access Journal,* Volume 6.

Leevy, J., Khoshgoftaar, T. & Bauder, R., 2018. A survey on addressing high-class imbalance in big data. *J Big Data,* 5(2).

Lee, Y., Han, D. & Ko, H., 2013. Reinforced AdaBoost Learning for Object Detection with Local Pattern Representations. *The Scientific World Journal,* Volume 2013, p. 14 pages.

Li, J., Chen, Q., Hu, X. & Yuan, p., 2021. Establishment of noninvasive diabetes risk prediction model based on tongue features and machine learning techniques. *International Journal of Medical Informatics,* Volume 149.

Lima, D., Sanches, R. & Pedrino, E., 2017. Neural Network Training Using Unscented and Extended Kalman Filter. *Robotics & Automation Engineering Journal,* 1(4).

Lin, F., 2008. Solving Multicollinearity in the Process of Fitting Regression Model Using the Nested Estimate Procedure. *Qual Quant ,* pp. 417-426.

Lin, J. & Wang, Y., 2011. Using a Novel AdaBoost Algorithm and Chous Pseudo Amino Acid Composition for Predicting Protein Subcellular. *Proten Pept Lett,* December, 18(12), pp. 1219-25.

Lin, K. & Yeh, C., 2012. Use of data mining techniques to detect medical fraud in health insurance in proceeding. *International journal of engineering and technology innovation,* 2(2), pp. 42-53.

Li, Q., Li, R., Ji, K. & Dai, W., 2015. *Kalman Filter and Its Application.* s.l., IEEE, pp. 74-77.

Li, S. & Zhang, Z., 2004. FloatBoost Learning and Statistical Face Detection. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE,* 26(9).

Li, S. Z., Zhang, Z., Shum, H.-Y. & Zhang, H., 2003. FloatBoost Learning for Classification. *Neural Information Processing Systems (NIPS),* Issue Advances in Neural Information Processing Systems 15.

Liu, Y., He, S., Ran, B. & Cheng, Y., 2018. A Progressive Extended Kalman Filter Method for Freeway Traffic State Estimation Integrating Multisource Data. *Wireless Communications and Mobile Computing,* Volume 2018.

Livieris, I. & et, a., 2020. Ensemble Deep Learning Models for Forecasting Cryptocurrency Time-Series. *MDPI - Algorithims.*

Li, V., Mitchell, R. & Boyle, B., 2016. The Divergent Effects of Transformational Leadership on Individual and Team Innovation. *Group & organization management,* 41(1), pp. 66-97.

Li, W., Wang, W. & Huo, W., 2020. RegBoost: a gradient boosted multivariate regression algorithm. *International Journal of Crowd Science,* 4(1), pp. 60-72.

Love, P. E. D. et al., 2018. Reduce rework, improve safety: an empirical inquiry into the precursors to error in construction. *Production Planning and Control,* 29(7).

Love, P. E. D. et al., 2018. Reduce rework, improve safety: an empirical inquiry into the precursors to error in construction. *Production Planning & Control,* 5(353–366), p. 29.

Luque, E., Carrasco, A., Martin, A. & Hera, A. d.-l., 2019. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition,* Volume 91, pp. 217-231.

Lynam, A., Dennis, J., Owen, K. & al., e., 2020. Logistic regression has similar performance to optimised machine learning algorithms in a clinical setting: application to the discrimination between type 1 and type 2 diabetes in young adults. *Diagn Progn Re,* 4(6).

Maclaurin, D., Duvenaud, D. & Adams, R., 2015. *Early Stopping is Nonparametric Variational Inference,* s.l.: arXiv:1504.01344v1 [stat.ML].

Mai, X., Lioa, Z. & Couillet, R., 2019. *A Large Scale Analysis of Logistic Regression: Asymptotic Performance and New Insights.* s.l., IEEE, pp. 3357-3361.

Malathi, A., and Baboo, S. & (2011b), 2011b. Evolving Data Mining Algorithms on the prevailing crime trend – An intelligence crime prediction model. *International Journal of,* Issue 6.

Malathi, A. & Baboo, S., 2011a. Algorithmic Crime Prediction Model Based on the Analysis of Crime Clusters. *Global Journal of Computer Science and Technology,* 11(11).

Malathi, P. & Suresh, G., 2014. *Recognition of isolated words of esophageal speech using GMM and gradient descent RBF networks.* s.l., IEEE, pp. 174 - 177.

Mandrekar, J., 2010. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *Journal of Thoracic Oncology,* 5(9), pp. 1315-1316.

Mangiafico, S., 2015. *An R Companion for the Handbook of Biological Statistics.* 1.09j ed. New Brunswick, NJ: Rutgers Cooperative Extension.

Mariani, S. & Ghisi, A., 2007. Unscented Kalman filtering for nonlinear structural dynamics. *Unscented Kalman filtering for nonlinear structural dynamics.,* Volume 49, pp. 131-150.

Marsland, S., 2009. *Machine Learning: An Algorithmic Perspective.* Second Edition ed. s.l.:Chapman & Hall/CRC.

MBASkills., 2021. *Curve (the values on each axis?).* [Online]
Available at: http://mbaskills.in/what-is-auc-roc-curve/
[Accessed 30 August 2021].

McDonald, J., 2014. *Handbook of biological statistics.* 3rd ed. Maryland: Sparky House Publishing, Baltimore.

McGinley, B. et al., 2010. SPIKING NEURAL NETWORKS FOR BREAST CANCER CLASSIFICATION USING RADAR TARGET SIGNATURES. *Progress In Electromagnetics Research C,* Volume 17, pp. 74-94.

Mellor, A. & Boukir, S., 2017. Exploring diversity in ensemble classification: Applications in large area land cover mapping. *ISPRS Journal of Photogrammetry and Remote Sensing,* Volume 129, pp. 151-161.

Meng, F. et al., 2016. The influence of organizational culture on talent management: A case study of a real estate company. *Journal of Chinese Human Resource Management,* 7(2), pp. 129-146.

Merjildo, D. & Ling, L., 2012. Enhancing the Performance of AdaBoost Algorithms by Introducing a Frequency Counting Factor for Weight Distribution Updating. *LNCS 7441,* pp. 527-534.

Miah, M., 2011. *Survey of Data Mining Methods in Emergency Evacuation Planning.* Wilmington North Carolina, USA, s.n.

Mishra, D., Das, A. & and Mishra, M., 2010. Predictive Data Mining: Promising Future and Applications. *Int. J. of Computer and Communication Technology,* 2(1).

Moody, J. & Darken, C., 1989. Fast learning in networks of locally-tuned processing units. *Neural Computation,* i(2), pp. 282-295.

Murphy, K. P., 2012. *Machine Learning: A Probabilistic Perspective (Adaptive Computation and Machine Learning series).* 1st ed. Machachutes: Machachutes institute of technology, The MIT Press.

Nabney, I., 2002. *NETLAB Algorithms for Pattern Recognition.* London: Springer.

Natekin, A. & Knoll, A., 2013. Gradient boosting machines, a tutorial. *Front Neurorobot,* 7(21).

Nie, F., Hu, Z. & Li, X., 2018. An investigation for loss functions widely used in machine learning. *Commincations in Information and Systems,* 18(1), pp. 37-52.

Nielsen, M. .., 2015. *Neural networks and deep learning.* s.l.:Determination Press.

NIH, 2020. *National Cancer Institute.* [Online]
Available at: https://bcrisktool.cancer.gov/calculator.html
[Accessed 13 March 2921].

Niu, Y. & Hu, L., 2016. *An Extended Kalman Filter Application on Moving Object Tracking.* s.l., s.n., pp. 1261-1268.

Nyce, C., 2007. *Predictive Analytics White Paper,* Malvern: American Institute for Chartered Property Casualty Underwriters/Insurance Institute of America.

Oladapo, V., 2014. The impact of talent management on retention. *Journal of business studies quarterly,* 5(3), p. 19.

Olga, M. & Steve, B., 2016. Implementing operations strategy through Lean processes within health care: The example of NHS in the UK. *International journal of operations & production management,* 36(11), pp. 1435-1457.

Oliveira, M., 2012. An Application of Neural Networks Trained with Kalman Filter Variants (EKF and UKF) to Heteroscedastic Time Series Forecasting. *Applied Mathematical Sciences,* 6(74), pp. 3675 - 3686.

Olsson, B., Laurio, ,. K. & Gudjonsson, L., 2001. A hybrid method for protein sequence modeling with improved accuracy. *Information science,* 139(1-2), pp. 113-138.

Pak, F., Kanan, H. & Alikhassi, A., 2015. Breast cancer detection and classification in digitalmammography based on Non-SubsampledContourlet Transform (NSCT) and Super Resolution. *Co m p u t e r Me t h o d s a n d P r o g r a m s i n B i om e d i c i n e,* Volume 122, pp. 89-107.

Park, C. & Cho, S.-B., 2003. *Evolutionary ensemble classifier for lymphoma and colon cancer classification.* s.l., s.n., pp. 2378-2385.

Parker, R. & Bradley, L., 2000. Organizational culture in the public sector: Evidence from six organizations. *International Journal of Public Sector Management,* 13(2), pp. 125-141.

Parliament, 2020. *Mental health impacts of COVID-19 on NHS healthcare staff.* [Online] Available at: https://post.parliament.uk/mental-health-impacts-of-covid-19-on-nhs-

healthcare-staff/

[Accessed 10 December 2021].

Patrick, J., 2017. *Machine Learning Algorithm Taxonomy.* [Online]

Available at: https://www.deepmarketer.com/blog/2017/1/30/machine-learning-algorithm-taxonomy

[Accessed July 28 2021].

Peng, D., Gu, F., Hu, H. X. & Liu, C., 2021. Addressing the multi-label imbalance for neural networks: An approach based on stratified mini-batches. *Neurocomputing,* Volume 435, pp. 91-102.

Philips., 2020. *Predictive analytics in healthcare: three real-world examples.* [Online]

Available at: https://www.philips.com/a-w/about/news/archive/features/20200604-predictive-analytics-in-healthcare-three-real-world-examples.html

[Accessed 25 May 2021].

Pintelas, P. & Livieris, I., 2020. Special Issue on Ensemble Learning and Applications. *MDPI-Algorithims.*

Polikar, R., 2006. Ensemble Based Systems in Decision Making. *IEEE Circuits and System magazine,* 6(3), pp. 21-45.

Polikar, R., 2012. Ensemble Machine Learning. In: C. Zhang & Y. Ma, eds. *Ensemble Machine Learning : Methods and Applications.* s.l.:Springer Science+Business Media, LLC, pp. 1-35.

Prechelt, L., 2012. Early Stopping - But When?. In: G. Montavon, O. G.B. & K. Müller, eds. *Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science.* s.l.:Springer, Berlin.

Qiao, X., Chang, W., Zhou, S. & Lu, X., 2016. *A prediction model of hard landing based on RBF neural network with K-means clustering algorithm.* s.l., IEEE, pp. 462-465.

RaEng., 2014. *and trust of the use of algorithms in decision-making,* London: Royal Academy of Engineering.

Ramadurai, S. et al., 2012. *Application of Unscented Kalman Filter to a Cable Driven Surgical Robot: A Simulation Study.* Saint Paul, s.n.

Ramdhony, A. & D'Annunzio-Green, N., 2018. A dialogic reframing of talent management as a lever for hospitableness. *Worldwide Hospitality and Tourism Themes,* 10(1), pp. 14-27.

Randhawa, K. et al., 2018. Credit Card Fraud Detection Using AdaBoost and Majority Voting. *IEEE Access,* Volume 6.

Raskutti, G., Wainwright, M. & Yu, B., 2014. Early Stopping and Non-parametric Regression: An Optimal Data-dependent Stopping Rule. *Journal of Machine Learning Research,* Volume 15, pp. 335-366.

Reeve, H.-J. & Brown, G., 2018. Diversity and degrees of freedom in regression ensembles. *Neurocomputing,* Volume 298, pp. 55-68.

Reeve, H. & Brown, G., 2018. Diversity and degrees of freedom in regression ensembles. *Neurocomputing,* Volume 298, pp. 55-68.

Reeves, S., Albert, M., Kuper, A. & Hodges, B., 2008. Qualitative research - Why use theories in qualitative research?. *BMJ Clinical Research ,* 337(7670).

Rhudy, M., 2015. *Increasing the Convergence Rate of the Extended Kalman Filter.* s.l., s.n.

Ribeiro, M., 2004. *Kalman and Extended Kalman Filters: Concept, Derivation and Properties,* s.l.: CiteSeer.

Rivals, I. & Personnaz, L., 1998. A recursive algorithm based on the extended Kalman filter for the training of feedforward neural models. *Neurocomputing,* 20(1-3), pp. 279-294.

Rodríguez-Jiménez, A., Becerra-Permuy, J., Bellas-Bouza, F. & Arias-Méndez, E., 2019. *Deep Learning application to learn models in Cognitive Robotics.* s.l., IEEE.

Rodríguez, J. & Maudes, J., 2008. Boosting recombined weak classifiers. *Pattern Recognition Letters,* 29(8), pp. 1049-1059.

Rokach, L., 2005. Ensemble Methods for Classifiers. In: *Data Mining and Knowledge Discovery Handbook.* s.l.:Springer US, pp. 980-1000.

Rokach, L., 2010. Ensemble-based classifiers. *Artificial Intelligence Review,* Volume 33, pp. 1-39.

Sagi, O. & Rockach, L., 2018. Ensemble learning: A survey. *WIREs Data Mining and Knowledge Discovery,* 8(4).

Saiti, ,. M. et al., 2020. Ensemble methods in combination with compartment models for blood glucose level prediction in type 1 diabetes mellitus. *Computer Methods and Programs in Biomedicine,* Volume 196.

Sapate, S. & Talbar, S., 2016. Mammograms, An Overview of Pectoral Muscle Extraction Algorithms Applied to Digital. *Studies in Computational Intelligence.*

Sarker, I., 2021. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN COMPUT. SCI.,* 2(160).

Sasaki, S., Comber, A. J., Suzuki, H. & Brunsdon, C., 2010. Using genetic algorithms to optimise current and future health planning-the example of ambulance locations. *International journal of health geographic,* 9(9).

Schapire, R., 1999. *A Brief Introduction to Boosting.* s.l., s.n.

Schapire, R., 2013. Explaining AdaBoost. In: *Empirical Inference.* s.l.:Springer, pp. 37-52.

Schapire, R. E., 2013. Explaining AdaBoost. In: B. L. Z. V. V. Schoelkopf, ed. *Empirical Inference Festschrift in Honor of Vladimir N. Vapnik.* s.l.:Springer, pp. 27-52.

Schapire, R. E. & Freund, Y., 2014. *Boosting - Foundations and Algorithms.* 1st ed. Cambridge, Massachusetts: The MIT Press.

Schapire, R. & Freund, Y., 2012. *Boosting: Foundations and Algorithms..* s.l.:MIT Press.

Schapire, R. & Freund, Y., 2014. *Boosting: Foundations and algorithms.* s.l.:MIT Press.

Schapire, R., Freund, Y., Bartlett, P. & Lee.W.S., 1998. Boosting the margin: A new explanation for the effectiveness of voting methods. *The Annals of Statistics,* 26(5), pp. 1651-1686.

Schapire, R. & Singer, Y., 1999. Improved boosting algorithms using confidence-rated predictions. *Machine Learning,* 37(3), pp. 297-336.

Schapire, R. & Singer, Y., 1999. Improved boosting algorithms using confidence-rated predictions. *Machine Learning,* 37(3), p. 297–336.

Schein, E., 2016. *Organizational Culture and Leadership.* 5th ed. San Francisco : The Jossey-Bass Business & Management Series.

Schein, P. & Schein, E., 2016. *Organizational Culture and Leadership.* 5th ed. s.l.:John Wiley & Sons, Incorporated.

Schmidt, J., Marques, M., Botti, S. & Marques, M., 2019. Recent advances and applications of machine learning in solid-state materials science. *npj Computational Materials ,* 5(83).

Schmidt, J., Marques, M., Botti, S. & Marques, M., 2019. Recent advances and applications of machine learning in solid-state materials science. *npj Computational Materials ,* 5(83).

Seiffert, C., Khoshgoftaar, T. M., Hulse, J. V. & Napolitano, A., 2008. *Resampling or Reweighting: A Comparison of Boosting Implementations.* s.l., s.n.

Seuwou, P. & Adegoke, V., 2021. The Changing Global Landscape With Emerging Technologies and Their Implications for Smart Societies. In: A. Nwajana & I. Ihianle, eds. *Handbook of Research on 5G Networks and Advancements in Computing, Electronics, and Electrical Engineering.* s.l.:IGI Global - Timely Knowledge.

Shahhosseini, M., Hu, G. & Pham, H., 2019. *Optimizing Ensemble Weights and Hyperparameters of Machine Learning Models for Regression Problems,* s.l.: Iowa State University Digital Repository.

Shareef, A., Zhu, Y., Musavi, M. & Shen, B., 2007. *Comparison of MLP Neural Networks and Kalman Filter for Localization in Wireless Sensor Networks.* Cambridge, MA, s.n.

Simon, D., 2002. Training Radial Basis Neural Networks with the Extended Kalman Filter. *Neurocomputing,* Volume 48, pp. 455-457.

Simon, D., 2013. *Using Nonlinear Kalman Filtering to Estimate Signals.* [Online] Available at: https://academic.csuohio.edu/simond/pubs/ESDNonlinear1.pdf [Accessed 06 February 2021].

Simsek, C. & Turk, K., 2016. Simplified Early Stopping Criterion for Belief-Propagation Polar Code Decoders. *IEEE Communications Letter,* 20(8), pp. 1515-1518.

Singh, S. & Thakral, S., 2018. *Using Data Mining Tools for Breast Cancer Prediction and Analysis.* Greater Noida, India, International Conference on Computing Communication and Automation (ICCCA).

Society, A. C., 2015. *American Cancer Society. Breast Cancer Facts & Figures 2015-2016.* [Online] Available at: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2015-2016.pdf [Accessed 05 May 2017].

Soni, R., Nigam, M. & Mitra, R., 2015. s.l., IEEE, pp. 178 - 183.

Souza, E. & Matwin, S., 2012. *Improvements to AdaBoost Dynamics.* Berlin, Heidelberg, Springer, pp. 293-298.

Sublime, J., Grozavu, N., Bennani, Y. & Cornuejols, A., 2015. *Collaborative Clustering with Heterogeneous Algorithms,* s.l.: IEEE.

Sun, Y., Todorovic, S. & Li, J., 2006. Reducing the overfitting of AdaBoost by controlling its data distribution skewness. *International Journal of Pattern Recognition and Artificial Intelligence,* 20(7), pp. 1093-1116.

Taylor, S., Levy, O., Boyacigiller, N. & Beechler, S., 2008. Employee commitment in MNCs: Impacts of organizational culture, HRM and top management orientations. *The International Journal of Human Resource Management,* 19(4).

Tewari, S. & Dwivedi, U., 2020. A comparative study of heterogeneous ensemble methods for the identification of geological lithofacies. *J Petrol Explor Prod Technol,* Volume 10, pp. 1849-1868.

Thomas, M., 2019. *15 Examples of Machine Learning in Healthcare That Are Revolutionizing Medicine.* [Online]
Available at: https://builtin.com/artificial-intelligence/machine-learning-healthcare
[Accessed 05 August 2021].

Thongkam, J., Xu, G. & Zhang, Y., 2008. *AdaBoost algorithm with random forests for predicting breast cancer survivability.* s.l., s.n., pp. 3062 - 3069.

Thunnissen, M., 2016. Talent management: For what, how and how well? An empirical exploration of talent management in practice. *Employee Relations,* 38(1), pp. 57-72.

Ting, K. & Witten, I., 1999. Issues in Stacked Generalization. *Journal Of Artificial Intelligence Research,* Volume 10, pp. 271-289.

Tsai, Y., 2011. Relationship between Organizational Culture, Leadership Behavior and Job Satisfaction. *BMC Health Services Research,* 11(98).

Tsoumakas, G., Partalas, I. & Vlahavas, I., 2008. *A Taxonomy and Short Review of Ensemble Selection.* s.l., s.n., pp. 41-46.

Ucar, A., Alpkan, L. & Elci, M., 2021. The Effect of Servant and Transformational Leadership Styles on Employee Creative Behavior: The Moderating Role of Authentic Leadership. *International Journal of Organizational Leadership,* 10(Special issue), pp. 99-119.

UCI, 2015. *UCI Machine Learning Repository.* [Online]
Available at: http://archive.ics.uci.edu/ml/
[Accessed 21 December 2015].

UK, C. R., 2018. *Cancer Research UK.* [Online]
Available at: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer
[Accessed 08 20 2018].

UmaMageswari, A., Ignatious, J. & Vinodha, R., 2012. A Comparitive Study Of Kalman Filter, Extended Kalman Filter And Unscented Kalman Filter For Harmonic Analysis Of The Non-Stationary Signals. *International Journal of Scientific & Engineering Research,* 3(7).

Vabalas, A., Gowen, E., E., P. & Casson, A., 2019. Recent advances and applications of machine learning in solid-state materials science. *PLoS one,* 14(11).

Valverde, J., 2018. *Difference between L1 and L2 regularization, implementation and visualization in Tensorflow.* [Online]
Available at: http://laid.delanover.com/difference-between-l1-and-l2-regularization-implementation-and-visualization-in-tensorflow/
[Accessed 02 August 2021].

Vapnik, V., 1995. *The Nature of Statistical Learning Theory.* s.l.:Springer.

Vašát, R., Kodešová, R. & Borůvka, L., 2017. Ensemble predictive model for more accurate soil organic carbon spectroscopic estimation. *Computers & Geosciences.*

Venkatesan, G. & Sasikala, V., 2019. A Statistical Analysis Of Migration Using Logistic Regression Model. *International Journal of Fcentific & Technology Research,* 8(10).

Vezhnevets, A. & Vezhnevets, V., 2005. *'Modest AdaBoost' – Teaching AdaBoost to Generalize Better,* s.l.: Graphicon.

Viola, P. & Jones, M. J., 2004. Robust Real-Time Face Detection. *International Journal of Computer Vision,* 57(2), pp. 137-154.

Vuttipittayamongkol, ,. P., Elyan, E. & Petrovski, A., 2021. On the class overlap problem in imbalanced data classification. *Knowledge-Based Systems,* Volume 212, p. 106631.

Walker, S., Khan, W., Katic, K. W. & Zeiler, W., 2020. Accuracy of different machine learning algorithms and added-value of predicting aggregated-level energy performance of commercial buildings. *Energy and Buildings,* Volume 209.

Wan, E. & Merwe, R. D., 2002. *The Unscented Kalman Filter for Nonlinear Estimation.* Lake Louise, Alberta, Canada, IEEE.

Wan, E. & Van Der Merwe, R., 2000. The Unscented Kalman Filter for Nonlinear Estimation, Communications, and Control Symposium. *Adaptive Systems for Signal Processing,* pp. 153-158.

Wang, J. et al., 2006. *Training RBF Networks with an Extended Kalman Filter Optimized Using Fuzzy Logic.* s.l., IFIP International Federation for Information Processing, pp. 317-326.

Wang, K., Wang, Y., Zhao, Q. & Meng, D., 2015. SPLBoost: An Improved Robust BoostingAlgorithm Based on Self-paced Learning. *JOURNAL OF LATEX CLASS FILES,* 14(8).

Wang, P., 2017. *The Application of Radial Basis Function (RBF) Neural Network for Mechanical Fault Diagnosis of Gearbox.* s.l., IOP Publishing Ltd.

Wang, Q. et al., 2017. A Novel Ensemble Method for Imbalanced Data Learning: Bagging of Extrapolation-SMOTE SVM. *Computational Intelligence and Neuroscience,* p. 11 pages.

Wang, S. et al., 2011. *RBF neural network parameters optimization based on paddy field algorithm.* s.l., IEEE, pp. 349-353.

Wang, W., 2008. *Some fundamental issues in ensemble methods.* Hong Kong, IEEE.

Wang, W., 2008. *Some fundamental issues in ensemble methods.* s.l., IEEE.

Wang, Y. et al., 2015. *Fuzzy radial basis function neural network PID control system for a quadrotor UAV based on particle swarm optimization.* s.l., IEEE, pp. 2580 - 2585.

Wankhade, P., 2010. *Cultural characteristics in the ambulance service and its relationship with organisational performance: evidence from the UK.* [Online]
Available at:
https://www.academia.edu/920660/Cultural_characteristics_in_the_ambulance_service_and_its_relationship_with_organisational_performance_evidence_from_the_UK
[Accessed 12 December 2021].

Warmuth, M., Glocer, K. & Vishwanathan, S., 2008. *Entropy regularized LPBoost.* Budapest, Hungary, s.n., p. 256–271.

Weedon-Fekjær, H., Romundstad, P. & Vatten, L., 2014. Modern mammography screening and breast cancermortality: population study. *MMJ,* Volume 348:g3701.

Wei, Y., Yang, F. & Wainwright, M., 2019. Early Stopping for Kernel Boosting Algorithms: A General Analysis With Localized Complexities. *IEEE Transactions on Information Theory,* 65(10), pp. 6685-6703.

Wei, Y., Yang, F. & Wainwright, M., 2019. Early Stopping for Kernel Boosting Algorithms: A General Analysis With Localized Complexities. *IEEE Transactions on Information Theory,* 65(10), pp. 6685-6703.

Whalen, S. & Pandey, G., 2013. A Comparative Analysis of Ensemble Classifiers: Case Studies in Genomics.

WHO, 2016. *GLOBAL REPORT ON DIABETES- World Health Organization,* Geneva 27, Switzerland: WHO Library Cataloguing-in-Publication Data.

WHO, 2017. *Breast cancer: prevention and control,* s.l.: World Health Organization.

WHOa, 2021. *Breast cancer now most common form of cancer: WHO taking action.* [Online]
Available at: https://www.who.int/news/item/03-02-2021-breast-cancer-now-most-common-form-of-cancer-who-taking-action
[Accessed 13 March 2021].

WHOb, 2020. *Diabetes.* [Online]
Available at: https://www.who.int/news-room/fact-sheets/detail/diabetes
[Accessed 13 March 2021].

Williams, N. & Beidas, J., 2018. Navigating the storm: How proficient organizational culture promotes clinician retention in the shift to evidence-based practice. *PloS one,* 13(12).

Wilson, L. F., 1996. *Analysis of Algorithmic Structures with Heterogeneous Tasks,* Hampton: IEEE.

Witten, I., Frank, E. & Hall, M., 2011. *Data Mining Practical Machine Learning Tools and Techniques.* 3rd ed. San Francisco(California): Morgan Kaufmann Publishers.

Wolpert, D., 1992. Stacked generalization. *Neural networks ,* 5(2), pp. 241-259.

Wsws, 2019. *UK ambulance staff quitting in record numbers.* [Online]
Available at: https://www.wsws.org/en/articles/2019/05/29/ambu-m29.html
[Accessed 12 December 2021].

Wu, C., Fan, W. & Liu, T., 2019. A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset. *Artificial Intelligence in Medicine,* Volume 10723.

Wu, S. & Nagahashi, H., 2014. *A new method for solving overfitting problem of gentle AdaBoost.* Hong Kong, s.n.

Wu, S. & Nagahashi, H., 2014. Parameterized AdaBoost: Introducing a Parameter to Speed Up the Training of Real AdaBoost. *Signal Processing Letters,* 21(6), pp. 687-691.

Wu, S. & Nagahashi, H., 2015. Analysis of Generalization Ability for Different AdaBoost Variants Based on Classification and Regression Trees. *Journal of Electrical and Computer Engineering.*

Wu, S. & Nagahashi, H., 2015. Penalized AdaBoost: Improving the Generalization Error of Gentle AdaBoost through a Margin Distribution. *Trans. Inf. & Syst.,* E98(11).

Wu, X.-x. & Liu, J.-u., 2009. *A New Early Stopping Algorithm for Improving Neural Network Generalization.* Changsha, China, 2009 Second International Conference on Intelligent Computation Technology and Automation.

Xiao, Z., Luo, Z., Zhong, B. & Dang, X., 2018. IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS. *Robust and Efficient Boosting Method Using the Conditional Risk,* 29(17).

Xie, W., Li, Y. & Ma, Y., 2015. Breast mass classification in digital mammography based on extreme learning machine. *Neurocomputing,* 73(3), pp. 930-941.

Xie, Y., Zhu, C., Lu, Y. & Zhu, Z., 2019. Towards Optimization of Boosting Models for Formation Lithology Identification. *Mathematical Problems in Engineering.*

Xing, Z., Guo, H., Dong, S. & Fu, Q. L. J., 2015. RBF Neural Network Based on K-means Algorithm with Density Parameter and Its Application to the Rainfall Forecasting. In: L. D. & C. Y., eds. *Computer and Computing Technologies in Agriculture VIII. CCTA 2014. IFIP Advances in Information and Communication Technology.* Cham: Springer.

Xu, L., Zhou, Z.-H. & Kwot, J., 2015. *Machine LearninSpringer Handbook of Computational Intelligence.* s.l.:Springer Nurture.

Yamashita, K., Chakraborty, G., Mabuchi, H. & https://ieeexplore.ieee.org/document/5578174/citations?tabFilter=papers#:~:text=Mabuch

i%20and%20M.-, M.-%. M., 2010. *An Efficient Method to Set RBF Network Paramters Based on SOM Training.* s.l., IEEE , pp. 426-431.

Yang, C.-H., Lin, Y.-U., Chuang, L.-Y. & Chang, H.-W., 2013. Evaluation of Breast Cancer Susceptibility Using Improved Genetic Algorithms to Generate Genotype SNP Barcodes. *IEEE/ACM Transactions on Computational Biology and Bioinformatics,* 10(2), pp. 361-371.

Yang, C., Shi, W. & Chen, W., 2016. Comparison of Unscented and Extended Kalman Filters with Application in Vehicle Navigation. *Journal of navigation,* 70(2), pp. 411-431.

Yang, H., Luo, Y., Ren, X. & Wu, M., 2021. Risk Prediction of Diabetes: Big data mining with fusion of multifarious physical examination indicators. *Risk Prediction of Diabetes: Big data mining with fusion of multifarious physical examination indicators.*

Yang, L., 2011. Classifiers selection for ensemble learning based on accuracy and diversity. *Procedia Engineering,* Volume 15, pp. 4266-4270.

Yeh, J., Wu, T. & Tsao, C., 2012. A Decision Support System for Predicting Hospitalization of Haemodialysis Patients. *International Journal of Biological and Life Sciences ,* 8(8).

Ying, X., 2019. An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series,* 1168(2).

Youlden, D., S.M.Cramb, Yip, C. & Baade, P., 2014. Incidence and mortality of female breast cancer in the Asia-Pacific region. *Cancer Biol Med,* Volume 11, pp. 101-115.

Yousefi, M. et al., 2018. Chaotic genetic algorithm and Adaboost ensemble metamodeling approach for optimum resource planning in emergency departments. *Artificial Intelligence in Medicine,* pp. 23-33.

Yu, T. & Zhu, H., 2020. Hyper-Parameter Optimization: A Review of Algorithms and Applications. *arxiv.*

Zain, M. & Tokhi, M., 2004. Improving performance in single-link flexible manipulator using hybrid learning control. *Jurnal Mekanikal,* December, Volume 18, pp. 13-28.

Zhang, H., Zhang, L. & Jiang, Y., 2019. *Overfitting and Underfitting Analysis for Deep Learning Based End-to-end Communication Systems.* s.l., s.n., pp. 1-6.

Zhang, M. et al., 2019. A literature review of failure prediction and analysis methods for composite high-pressure hydrogen storage tanks. *International Journal of Hydrogen Energy,* 44(47), pp. 25777-25799.

Zhang, T., 1999. *Analysis of Regularized Linear Functions for Classification Problems,* s.l.: s.n.

Zhang, Y. & Yang, H., 2013. *An Application of Radial Basis Function Network and Genetic Algorithm to Fashion Design System.* s.l., IEEE.

Zhang, Y.-y., Zhou, X.-b., Wang, Q.-z. & Zhu, X.-y., 2017. Quality of reporting of multivariable logistic regression models in Chinese clinical medical journals. *Medicine,* 96(21).

Zhang, Z. & Z., W., 2017. *Advanced statistics using R.* Granger: ISDSA Press.

Zheng, T. et al., 2017. A machine learning-based framework to identify type 2 diabetes through electronic health records. *International Journal of Medical Informatics,* Volume 97, pp. Pages 120-127.

Zhou, Z.-H., 2012. *Ensemble Methods Foundations and Algorithms.* s.l.:Chapman and Hall/CRC Press.

Zhou, Z.-H., 2012. *Ensemble Methods Foundations and Algorithms.* London: CRC Press Taylor & Francis Group, LLC.

Zhou, Z., 2012. *Ensemble methods: foundations and algorithms.* Boc Raton, FL: Chapman & Hall/CRC.

Zhukova, A. et al., 2017. Ensemble methods of classification for power systems security assessment. *Applied Computing and Informatics.*

Zu, S., Yin, S., Srinivasan, R. & Helander, M., 2012. Proactive Alarms Monitoring using Predictive Technologies,. In: I. Karimi & R. Srinivasan, eds. s.l.:Elsevier, pp. 1537-1541.