FIRAT UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
TÜRKİYE

HEART SOUNDS CLASSIFICATION USING DEEP LEARNING ALGORITHMS

Mohammed Mansur ABUBAKAR

Master's Thesis

DEPARTMENT OF COMPUTER ENGINEERING

MAY 2022

FIRAT UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
TÜRKİYE

Department of Computer Engineering

Master's Thesis

HEART SOUNDS CLASSIFICATION USING DEEP LEARNING ALGORITHMS

Author
Mohammed Mansur ABUBAKAR

Supervisor
Assoc. Prof. Dr. Taner TUNCER

MAY 2022
ELAZIG

FIRAT UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
TÜRKİYE

Department of Computer Engineering

Master's Thesis

Title:	Heart Sounds Classification Using Deep Learning Algorithms
Author:	Mohammed Mansur ABUBAKAR
Submission Date:	10 March 2022
Defense Date:	11 May 2022

THESIS APPROVAL

This thesis, which was prepared according to the thesis writing rules of the Graduate School of Natural and Applied Sciences, Fırat University, was evaluated by the committee members who have signed the following signatures and was unanimously approved after the defense exam made open to the academic audience.

		Signature
Supervisor:	Assoc. Prof. Dr. Taner TUNCER Firat University, Faculty of Engineering	Approved
Chair:	Assoc. Prof. Dr. Ebubekir ERDEM Firat University, Faculty of Engineering	Approved
Member:	Asst. Prof. Dr. İbrahim Rıza HALLAÇ Alanya Alaaddin Keykubat University, Faculty of Engineering	Approved

This thesis was approved by the Administrative Board of the Graduate School on

.... / .... / 20 ....

Signature

Prof. Dr. Kürşat Esat ALYAMAÇ
Director of the Graduate School

DECLARATION

I hereby declare that I wrote this Master's Thesis titled “Heart Sounds Classification Using Deep Learning Algorithms” in consistent with the thesis writing guide of the Graduate School of Natural and Applied Sciences, Firat University. I also declare that all information in it is correct, that I acted according to scientific ethics in producing and presenting the findings, cited all the references I used, express all institutions or organizations or persons who supported the thesis financially. I have never used the data and information I provide here in order to get a degree in any way.

11 May 2022

Mohammed Mansur ABUBAKAR

PREFACE

The purpose of this master's thesis is to introduce and show some practical applications of deep learning in medical applications; cardiovascular diseases diagnosis to be precise. Deep learning consists of various applied methods. Some limiting factors of the seminar study include accessibility to real patient medical data for public use. This study should be of interest to medical researchers, academicians, health practitioners, AI/deep learning researchers from both public and private institutions. It should also be of interest to scholars of deep learning and medical sciences. This thesis is written to meet the Graduate School of Natural and Applied Sciences requirements of the Computer Engineering master's program at Firat University. This thesis study was conducted from January 2021 to February 2022.

The research title was formulated together with my supervisor, Assoc Prof. Dr. Taner Tuncer. The research work started off difficult. However, while conducting the extensive study I've gotten to solve the problems stated. I was lucky enough; my supervisor was always available and willing to assist my efforts.

I would like to thank my supervisor for his remarkable guidance and support during this entire process. I also wish to thank all the authors and publishers of the works cited. Without their immense contribution to the field, I wouldn't have been able to conduct this study.

I would like to extend my profound gratitude to the Nigerian National Information Technology Development Agency (NITDA) for granting me a scholarship to study this masters' program.

Mohammed Mansur ABUBAKAR
ELAZIG, 2022

TABLE OF CONTENTS

Page

ABSTRACT

Heart Sounds Classification Using Deep Learning Algorithms

Mohammed Mansur ABUBAKAR

Master's Thesis

FIRAT UNIVERSITY
Graduate School of Natural and Applied Sciences

Department of Computer Engineering

May 2022, Page: xi + 56

The ever-evolving medical field driven by the applications of deep learning is changing the practice in many ways. Heart sounds are critical aspects that have to do with cardiac disease diagnosis. These are techniques in which heart abnormalities can be detected using signals gotten from heart sounds. The eminence of heart disease in the world today makes a cardiologist a very important health professional. Likewise, the role of artificial intelligence in the treatment of the cardiovascular system - which includes the heart, the blood vessels, etc. Such systems assist cardiologists in promoting heart health in patients. The classic diagnostic method is through cardiac auscultation to detect abnormalities in heart sounds. Thus, the accuracy of the heart auscultation is highly important in the diagnostic process to screen outpatients with or without heart diseases. This made it obvious there is a need for a more accurate and precise computer-aided diagnosis (CAD) system to further assist cardiologists in the field by examining heart sounds to provide a detailed outline. In this study, deep learning algorithms are implemented for the detection of abnormal and normal heart sounds using classification techniques. This study outlines the overview of deep learning-based heart diseases diagnosis in the context of experimental research. Furthermore, this study highlights research work that indicates the advances in the application of deep learning technology in medical sciences.

Keywords: CVDs, CNN, Deep Learning, Heart Sounds Classification, Medical Signals, Relief.

ÖZET

Derin Öğrenme Algoritmaları Kullanarak Kalp Sesleri Sınıflandırması

Mohammed Mansur ABUBAKAR

Yüksek Lisans Tezi

FIRAT ÜNİVERSİTESİ
Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

Mayıs 2022, Sayfa: xi + 56

Derin öğrenme uygulamalarının yönlendirdiği sürekli gelişen tıp alanı, uygulamayı birçok yönden değiştiriyor. Kalp sesleri, kalp hastalığı teşhisi ile ilgili kritik unsurlardır. Bunlar, kalp seslerinden alınan sinyaller kullanılarak kalp anormalliklerinin tespit edilebildiği tekniklerdir. Bugün dünyada kalp hastalığının saygınlığı, bir kardiyologu çok önemli bir sağlık profesyoneli yapmaktadır. Aynı şekilde, kalp, kan damarlarını vb. içeren kardiyovasküler sistemin tedavisinde yapay zekanın rolü. Bu tür sistemler kardiyologlara hastalarda kalp sağlığını geliştirmede yardımcı olur. Klasik tanı yöntemi, kalp seslerindeki anormallikleri tespit etmek için kalp oskültasyonudur. Bu nedenle ayaktan kalp hastalığı olan veya olmayan hastaların taranması için kalp oskültasyonunun doğruluğu tanı sürecinde oldukça önemlidir. Bu, ayrıntılı bir taslak sağlamak için kalp seslerini inceleyerek sahadaki kardiyologlara daha fazla yardımcı olmak için daha doğru ve kesin bir bilgisayar destekli tanı (CAD) sistemine ihtiyaç olduğunu açıkça ortaya koydu. Bu çalışmada, sınıflandırma teknikleri kullanılarak anormal ve normal kalp seslerinin tespiti için derin öğrenme algoritmaları uygulanmıştır. Bu çalışma, deneysel araştırmalar bağlamında derin öğrenmeye dayalı kalp hastalıkları teşhisine genel bir bakış sunmaktadır. Ayrıca, bu çalışma, derin öğrenme teknolojisinin tıp bilimlerinde uygulanmasındaki ilerlemeleri gösteren araştırma çalışmalarını vurgulamaktadır.

Anahtar Kelimeler: CVD'ler, CNN, Derin Öğrenme, Kalp Sesleri Sınıflandırması, Tıbbi Sinyaller, Rahatlama.

LIST OF FIGURES

Page

LIST OF TABLES

LIST OF APPENDICES

Page

ABBREVIATIONS

Abbreviations	
AI	: Artificial Intelligence
ANN	: Artificial Neural Networks
AUC	: Area Under the Curve
CADS	: Computer-Aided Diagnosis Systems
CNN	: Convolutional Neural Network
CRP	: Corneal-Retinal Potential
CSA	: Conventional Statistical Analysis
CT	: Computed Tomography
DL	: Deep Learning
DOI	: Depth of Interaction
DT	: Decision Tree
ECG	: Electrocardiography
FC	: Fully connected
FN	: False Negatives
FP	: False Positives
kNN	: k-Nearest Neighbor
LS	: Laplacian Score
LPP	: Locality Preserving Power
LSTM	: Long Short-Term Memory
MFCC	: Mel-Frequency Cepstrum Coefficient
MIT	: Medical Imaging Techniques
MIQ	: Mutual Information Quotient
ML	: Machine Learning
MRI	: Magnetic Resonance Imaging
MRMR	: Minimum Redundancy Maximum Relevance
NCA	: Neighborhood Component Analysis
NN	: Neural Networks
PCG	: Phonocardiogram
PLSR	: Partial Least Squares Regression
RF	: Radio Frequency
RNN	: Recurrent Neural Networks
ReLU	: Rectified Linear Units
STEM	: Science, Technology, Engineering and Mathematics
STFT	: Short Time Fourier Transform
TN	: True Negatives
TP	: True Positives
VHD	: Valvular Heart Diseases
WT	: Wavelet Transform
X-Ray	: X-Radiation

1. INTRODUCTION

Artificial Intelligence has been an effective way for machines to replicate human-level intelligence, i.e., learning and reasoning in the machine. This remarkable evolution is rapidly changing the field of medical diagnosis. The field of AI as a research area is often divided into more specific research sub-areas such as Applied deep learning in science [1], Digital Image Processing [2] et al. AI has captured the interest of many researchers ever since its emergence, it is a trendy area of technology with the most attention from researchers today. Alongside AI comes machine learning, a subset of AI often built on statistical techniques that enable machines to improve performance through learning and experience. Historical data is used as input in such algorithms to predict new output values. It is considered a broad area of research, with a lot of methods that have been developed, such as classification, clustering, neural networks, regression, decision tree, deep learning, et al. [3] [4]

Machine learning has grown at a faster rate in recent years than it had previously. DL is one of the most widely adopted ML algorithms. This has attracted much attention, especially in computer-aided diagnosis, because it often yields much higher accuracy, which sometimes goes beyond human perception [5]. The exceptional advancements in DL have a vast potential for medical applications [6]. Here we simulate the functionality of our brain cells called neurons in depth. This concept gave birth to the idea of neural networks. There has been so much improvement in medical data analysis using machine learning algorithms, most especially applying neural network techniques. The attention DL gets today creates an opportunity for many new possibilities applying deep learning in disease diagnosis. Although deep learning is often used in several other domains, our area of interest is medical signal processing and predictive diagnosis et al. [7], [8]. The medical field is widely considered one of the most critical areas where machine learning algorithms are applied. ML application spans from the effective use of medical imaging techniques employed in creating visual representations of the inner parts of a body for clinical analysis and medical intervention etc. ML enables computers to learn by using many descriptions in the same way that people learn. Using DL, computers learn to study and perform classification tasks on medical signals, which are recordings of an organism's biological effects, ranging from RNA (ribosome) and protein structures to brain and heart cycles, which are commonly used in the diagnosis of medical conditions like heart diseases, brain disorders, and epilepsy or some other kind of disorder.

With the emergence of ML and DL algorithms, the usage of AI has changed considerably during the last few decades. Thus, expanding the applications of AI to create opportunities in fields such as engineering, technology, and medical sciences. However, AI in its early stages was primarily focused on developing machines to make decisions that only a human could make. This can be seen in developing the first industrial robot arm, which became a part of the sequences in manufacturing at a major American car manufacturers factory in 1961. This system basically carried out the task of automating the die casting at the plant. We also saw the Shakey, built in 1966, at “Stanford Research Institute”, world's first electronic human. With all these AI inventions in engineering and technology, the advancement of AI in medicine was slow until few years back, when data became extremely important in driving AI and its subdomains. The application in medicine includes but not limited to personalized health care systems, predictive models used in diagnosing diseases and their treatment, etc.

Furthermore, AI and its applications could potentially be used in preventative medicine [9]. This technology dramatically improves diagnostic accuracy, efficiency, and monitoring efforts to provide care in the health sector. This also leads to an excellent overall patient outcome that promotes the growth of AI research in bio medical domain.

The problems addressed are applying deep learning technologies in a more accurate diagnostics, prediction, and features extraction from medical data (sound, images, and signals), respectively. Deployment of such DL tools will assist cardiologists with the detection of abnormality from cardia auscultation recordings performed on patients. It will also assist researchers in discovering the hidden opportunities of health data and DL algorithms in computer-aided analysis of heart diseases accurately to help administer better treatments. Thus, addressing these problems using DL tools will result in a better robust medical system. Hence, these problems are among those that were unsolvable with just traditional ML algorithms like SVM, KNN, Decision tree among others. Thus, deploying DL algorithms in solving such problems in the medical domain are discussed in this study. These DL algorithms make use of neurons to boost computational labor and deliver accurate outcomes in problem solving.

Cardiac abnormalities are regarded as one of the life’s most threatening abnormalities today. Among the leading cause of death globally remains heart disease. Compared with several other causes of death, cardiovascular diseases often result in more people dying every year, as evidenced by WHO publications. The most frequent cause worldwide is cardiovascular (heart) illnesses. In 2016, 17.9 people worldwide were predicted to have died from cardiovascular (CVD). This accounts for roughly 31% of all mortality globally, with heart problems accounting for 85% of those deaths [10]. Most heartbeat abnormalities are reflected in heart sounds.

Thus, accuracy in cardiac auscultation is a crucial process in diagnosing heart conditions in patients; the benefits of addressing these problems using DL algorithms is to help both clinicians and researchers comprehend and analyze the disease better to provide patients with the best possible care efficiently and effectively. It also expands to using such algorithms for a non-invasive monitoring and evaluation of the condition and the effectiveness of the medical intervention provided. It also cuts down cure development costs, efforts, and time, which a typical example could be, the instance of the Capecitabine (also known as Xeloda) is a breast cancer medication. After some trial with only 162 subjects, it became licensed in 1998 due to vascular decrease on CT scans. These CT scans could be fed into a DL model for accurate classification using a features extraction mechanism in DL technology while cutting down the time it takes if the CT image was to be examined by a medical practitioner [11].

Some of the contributions of this thesis study to knowledge include matching or exceeding the desired performance of any proposed or existing computer-aided diagnosis system so that this study will help enhance the field of cardiac abnormalities identification at an early stage. This is achieved as an enhancement to any related work that fails to deliver a sound model with a much better performance when applied.

The goals of this study include: First, we will provide an overview of the algorithms applied. Second, we will discuss in-depth the adoption of DL techniques in computer-aided diagnosis systems. We will further discuss some experimental studies done in classifying heart sounds. We will see how DL is applied in the diagnostics and analysis of three classes of heart conditions: normal, murmur, and extrasystole. Thus, this study is structured in two major parts; where the first part introduces general concepts and principles of ML and DL, medical signals, selection, and classification algorithms. We will further discuss applying deep learning methods in ECG and scalograms (wavelet transform) adopted for this study. The second part provides a detailed discussion on the methodology used by the author(s), which entails the application of the proposed hybrid heart disease diagnosis model coupled with the performance results achieved and discussions drawn and concluded.

The aim of our study includes: (a) Introduction and discussion on deep learning in heart sound classification with reference to related studies; (b) Point out how deep learning has been applied in heart conditions, from medical diagnostics to analysis, and from classification to disease prediction; (c) Serve as an edge for cardiologists and researchers interested in experimenting and perhaps contributing to the field of deep learning for computer-aided diagnosis systems by identifying rich content, resources, interesting data sources and problems related to cardiovascular abnormalities. Furthermore, our study will conclude with limitations and future works on deep learning in heart conditions [12]

2. LITERATURE REVIEW

Different approaches have been applied on the popular Peter J Bentley – Heart sounds dataset, but the accuracy levels obtained from these approaches was because of longer computational processing time. Technology and medical science are extensively seen to be among the most important sectors adopting the wide and rapid advancement/growth today. Several technological advancements such as, computer-aided diagnosis of critical illnesses have been applied to the medical field to facilitate performance using data and information. Researchers identified that, for someone to fully understand the relation between technology and medical science, they must first and foremost understand what technologies are applied and used in medical sciences. For example, the applications of ML systems in bio-medical engineering, heart diagnosis systems etc.

Deep learning is an idea that emerged from the concept of neural network research, in which there are multiple layers models with several hidden layers. The application of this technology for medical purposes is however recent and still being explored. Nonetheless, researchers have been tirelessly improving this domain. Thus, we'll look at some of the approaches of several other researchers in this field.

Zeinali et al. (2022) proposed algorithm is used to diagnose heart diseases in heart sounds. The targets set for this paper is the effective identification and classification of heart sounds data accordingly. The proposed algorithm work by extracting the input data using signal processing techniques. Selection functions are applied for the purpose of getting best features out of the newly extracted feature, reducing the dimensionality while obtaining the optimal features for classification. However, the proposed method performed with accuracy level of 95% for multiclass classification problem and 98% accuracy score for binary classification problem. [13]

In 2021 [14], Deep learning approaches for heart audio categorization was indeed the title of the study project. The research of basic tenets and artificial intelligence utilized for ECG signals classifications is an important part of the research. The study also identifies key features used as input vectors, preprocessing methods, development of DL models used for patterns identification in cardiac rhythm.

In [15], researchers indicate that extracting feature vector from heart audio data can be done using components based on the wavelet transform (i.e., time and frequency domains). The study proposed an ideal segmentation algorithm to extract the heart sound components accordingly. Jia et al. [16] conducted a study that used actual data to predict the outcome with fuzzy network methods. The researchers collected the data used in the study themselves.

Utilizing new algorithms to classify heart sounds using local binary patterns coupled with deep learning is a method used by Mehmet et al. [17] get vector values from the audio recordings as input data. It feeds generated features into 1Dimension-CNN to successfully carry out final output. In this study, used two readily available datasets, PASCAL, and Physio Net. This experiment achieved 91.66% accuracy score were used to determine efficiency.

In [18] 2021 Alonso-Arevalo et al. propose a method that does segmentation of heart sounds using PCG signal. The proposed algorithm is further divided into detection phase and selection phase. A function indicating sound events is gotten from the computations of spectral flux (i.e., a measurement of how fast the spectra of such a signal change over time). Detection is done by carefully identifying and picking out time position, then selection is done by solving optimization problem. The method resulted in F1 score of 93.6% and 87.5% for the two variants; genetic algorithms (SGA) and the second is based on differential evolution (DE).

Gárate-Escamila et al. [19] worked towards reducing and understanding symptoms of heart diseases. The study aimed at proposing a solution towards dimensionality reduction in identifying heart diseases through the application of feature selection method.

In 2020 [20], Deng suggested an entirely different classification method which is unique from the conventional acoustic classification methods that are often insufficient in detecting and classifying heart sound. Such models result in a degraded performance because of their complex nature. Thus, to devise a higher performing model, the research team proposed a method that Mel-frequency cepstrum coefficient (MFCC) is increased, and multilayer Recurrent neural networks is used. This approach starts by computing the MFCC before even splitting the ECG signals wave. The dynamic features of the cardiac data are detailed using an image processing technique centered on MFCC. Lastly, CRNN accepts the features extracted by the MFCC. This is used for the feature learning and later perform the heartbeats classification. This designed DL technique uses the extracted features from the CNN and RNN.

Mohanad et al. proposed a method in 2020 which was able to conduct research to study efficient application of DL in diagnosing cardiac diseases. The proposed model was built on the combination of two types of networks, CNN and RNN, the model was a Bi-directional long short-term memory (BiLSTM), in other words a CNN-BiLSTM network. The proposed method attained 99.3% accuracy, sensitivity 98.3%, and specificity 99.6% respectively with and AUC of 0.998 [21].

Baghel et al. [22] in 2020, proposed a CNN-based model due to its high accuracy and robustness to detect CVDs efficiently and effectively from heart sounds. The proposed model employed some data augmentation techniques to improve the overall performance. After a test set up, the model achieved 98.6% accuracy, which diagnose several cardiac conditions.

In [23] 2019, Zhang et al. The researchers proposed a more effective and convenient way to diagnose heart diseases. A novel method that detects abnormalities in heart sounds using temporal features (spectrogram images) and LSTM. The model used the technique called "Fourier transform" to generate the temporal elements which were then used to calculate the magnitude difference in different frequency range. This is where the LSTM network came into play. The performance of the proposed model resulted in a 94.1% overall accuracy score.

Study by Hamidi et al. (2018) [24] proposed a method that eliminates expensive and time-consuming segmentation process in building the model. The team highlighted two different approaches to feature extraction when training the model. The first approach is using the technique "curve fitting" to extract the information apparent in events of heart audio recordings converted to wave signals. Hence, second proposed approach, fractal features are stacked to obtain the most powerful features using MFCC. The model was built for the purpose of differentiating pathologic heart sounds. This experiment applied 6 different datasets just as to assess the performance and compare with accurate models. Some of the classification algorithms employed in this study are kNN using Euclidean distance. It is concluded that the proposed algorithm resulted in a much better performing method in comparison to wavelet transform amongst others. The overall accuracy achieved on the proposed model is 91%.

In 2018, spectral features-based study was proposed by Bozkurt et al. [25] designing a solution that detects issues related to human heart. Audio recordings of the affected heart are the input to the proposed automatic CNN-based solution. This study uses PCG signal while considering features from (MFCC and Mel-Spectrogram). Upon conducting series of tests on two high quality databases, the proposed method gave 81.5% mean accuracy, 84.5% sensitivity and 78.5% specificity scores respectfully.

In 2017, Rubin et al. [26] presented a research work that applies deep learning techniques on heart sounds. By combining these methods, the model was able to classify heartbeat samples using machine learning approaches and Spectral domains: segmenting the initial basic heart sounds, applying some transformation techniques, training, and classification. This DL-based proposed model is used in cardiac auscultation i.e., used to recognize any form of abnormality in heart sounds. The algorithm presented employs key features like, time frequency and heat maps such as scalograms coupled with a CNN-based DL algorithm. Improving performance of the model requires this CNN-based architecture to be trained using a modified loss function to directly improve the effect it has on the model's results. They conducted testing on the built model in the 2016 PhysioNet cardiology contest, where it emerged the best model with a specificity score 95% and an overall score 84%.

Zhang and other researchers, [27] suggested in 2017 the classification of cardiac sounds using resized wavelet transforms with partial least squares regression. The researcher's approach to model building includes the following: estimating the cardiac rounds, resizing of the spectrogram, feature selection and output class prediction. The first part of the model building approach applies average with energy envelop precisely “Shannon”. followed by wavelet transforms used for generating features, where these wavelet transforms varied. Thus, obtaining the important set of vectors from the enhanced wavelet transform, Zhang et al. employed the PLSR as a feature selection technique which were then applied on the scaled spectrograms. This selection algorithm gave the model an edge on performance because it uses the category information during the dimension reduction process. Lastly, SVM classification algorithm was applied to obtain the output.

Rahhal et al. [28] in 2016, they suggested how we can go about improving the performance in which neural networks are used to extract and select the best possible features for outcome prediction. The trend in today's medical technology domain is deep learning for medical applications. This is a newly adopted technology that is currently undergoing thorough exploration that entails various research and practical application of DL techniques in the medical field. There are several studies that are related to heart sounds classification using deep learning methods for abnormality detection. Several of the papers referenced in this review section highlights the recent categorization objectives of such technologies. DL methods are frequently used on ECG data, the paper states Convolutional Networks (CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units as examples of deep neural networks (GRU).

In 2016 [29], Azmy proposed a new method for categorizing cardiac sounds. This method relies on statistical computation of the coefficients of a freshly developed Fourier transform to extract information from heart rhythm. The model further carries out classification using support vector machine (SVM). The researcher used 40 as the size of features with 90 heart sounds as training set, and 64 audio recordings of heartbeat as test data. The proposed method performed with results like 92% accuracy level, specificity of 95% and sensitivity of 90%. The research study further indicates how this newly developed tool can aid clinicians in diagnosing CVD in patients at an early stage.

Deng et al. (2016) suggested a new framework for automated heart sound echocardiography and monitoring of vital signs. This proposed model works without segmentation. In this model, the first part extracts the autocorrelation features of the heart signals which are gotten derived from the semi envelope derived from sub-band parameters These auto - correlation features are also combined to create integrated representations that incorporate diffusion maps. Lastly, merged features forwarded as input to the SVM classification algorithm for classifying the heart sounds as a final output. The proposed method achieved a best precision score of 91% which makes it outstanding in comparison with other models [30].

In 2015 [31], a multi-modal features technique was proposed by Randhawa & Singh for cardiac auscultation. PCG signal is used in this study to classify cardiac sounds into three categories: normal, systolic sound, and diastolic murmur. Feature extraction technique using spectrum analyzers to further reduce the dimensionality of the features creating fewer features that are more significant. This selection process removed the redundant features which were then used to classify the heart signals into the three output classes. Classifiers such as kNN, Fuzzy kNN alongside other models such as ANN are among the key methods employed in this research. The best performing models were the two out of three models with kNN and Fuzzy kNN as the classification algorithm. These models resulted in a higher accuracy level of 99.6%.

Varghees et al. [32], suggested robust approach on total variation filtering cardiac rhythm detection from medical-based signal. Validation strategy applied in the proposed model include the use of pruned heartbeats. The results achieved from the proposed model indicates how accurate the model identifies heart conditions present in the heart signal samples. The model achieved a 93.56% accuracy, with a feature suitable for real-time health monitoring technology.

In 2012, Safara et al. [33] conducted a heart sound analysis using wavelet transform with the entropy calculated for feature vectors. In this study, evaluating the discriminatory power of the generated features allowed for five classification tasks to be performed. However, this resulted in the best performance achieved using BayesNet with 96.94% accuracy. This conclusion draws attention onto how effective the proposed wavelet solution is used to identify patterns in heart rhythms.

New method employed in analyzing heart sound is proposed by Yuenyong et al. [34]. A difficult task in analyzing heart signals for diagnosis is segmentation of the data. This is because of noisy sounds like murmurs etc. The researchers took notes on varied heart rates utilizing information from the collected heart sounds and extracted an equivalent number of heartbeat rounds from the heart sounds. This approach consists of envelop identification, cardiac cycle length computation, extraction of features with Discrete Fourier Transform, feature selection using principal component analysis, and categorization with neural net bagging predictors. The procedure was put to the test using data from online libraries of heart recordings. The average classification performance was 92% in the case of data that is not noisy, while giving a 90% accuracy under noisy data. However, various experiments on the suggested technique have been carried out to solve any potential difficulties. The results show promising results on a dataset with a significantly larger size and range of heart sounds.

Ahlstrom et al. [35] proposes a model applying Pudil's progressive float forward selection (SFFS) method, a cross subset features were produced. In contrast to numerous different neural network classification techniques. This proposes a multi subset approach that performed the best, with 86 percent correctly predicted classification compared to 68 percent for the second-place finisher. A summary of related work introducing other researchers' works is seen in Table 2.1.

Table 2.1. Summary of related work.

S/N	Reference	Data size	Method	Results
1	Zeinali et al., 2022 [13]	650 sample	Signal processing algorithm	95% multiclass and 98% for binary classification problem.
2	Chen et al., 2021 [14]	N/A	CNN & RNN	97.63% accuracy
3	Mehmet et al., 2021 [17]	-	Dimensional Convolutional Neural Network (1D-CNN)	91.66% accuracy
4	Alonso-Arevalo et al., 2021 [18]	3,000 PCG signal	-	F1 score of 93.6% and 87.5% for the two variants
5	Deng et al., 2020 [20]	-	MFCC features and CRNN	98% accuracy
6	Gárate-Escamila et al., 2020 [19]	74	Chi-square test, Principal components analysis and random forests/Naïve Bayes/logistic regression/gradient-boosted tree/decision tree CHI-PCA-RF	Accuracy of 99.0%, precision score 100.0%, recall score of 96.8%, and F1 score 98.4%
7	Mohanad et al., 2020 [21]	9,600 samples	CNN-BiLSTM network	accuracy of 99.3%, sensitivity 98.3%, and specificity 99.6% and AUC of 0.998
8	Baghel et al., 2020 [22]	1000 audio samples	CNN	98.6% accuracy
9	Zhang et al., 2019 [23]	3153 records	LSTM	94.1% accuracy
10	Hamidi et al., 2018 [24]	3126 heart sound signals	Curve fitting method MFCC	91% accuracy
11	Bozkurt et al., 2018 [25]	83 PCG samples and 336 recordings	CNN	81.5% mean accuracy, 84.5% sensitivity and 78.5% specificity
12	Rubin et al., 2017 [26]	4,430 recordings	CNN architecture with MFCC heat maps	Specificity score 95%

Table 2.1. (Continued).

S/N	Reference	Data size	Method	Results
13	Zhang et al., 2017 [27]	683 records	Scaled spectrogram + PLSR + SVM	90% specificity
14	Rahhal et al., 2016 [28]	-	DNN such as CNN, LSTM, GRU	-
15	Azmy, 2016 [29]	154 heart sounds	DWT + SVM	Accuracy 92%, specificity 95% and sensitivity 90%
16	Deng et al. 2016 [30]	Dataset-A contains 176 records Dataset-B contains 656 records	SVM - Diffusion maps (DM)	Precision score 91%
17	Randhawa et al. 2015 [31]	144 samples	kNN, Fuzzy kNN, ANN and Naïve Bayes	99.6% accuracy
18	Varghees et al., 2014 [32]	-	Total variation filter (TVF) with defined low frequency using low-pass filtering (LPF)	93.56% Accuracy for TVF and 91% for LPF
19	Castro et al., 2013 [15]	-	Singular value decomposition (SVD) and energy envelope peaks	92.1% accuracy
20	Safara et al., 2012 [33]	350 heart sounds	Bayes Net and SVM, wavelet packet entropy	96.94% accuracy
21	Yuenyong et al., 2011 [34]	-	Envelope detection, cardiac cycle calculation, features extraction using DWT, PCA selection, and classification using neural network	92% accuracy not noisy data, 90% accuracy in noisy data
22	Ahlstrom et al., 2006 [35]	445 cardiac cycles collected from 36 people	neural network and sequential floating forward selection (SFFS)	86% accuracy

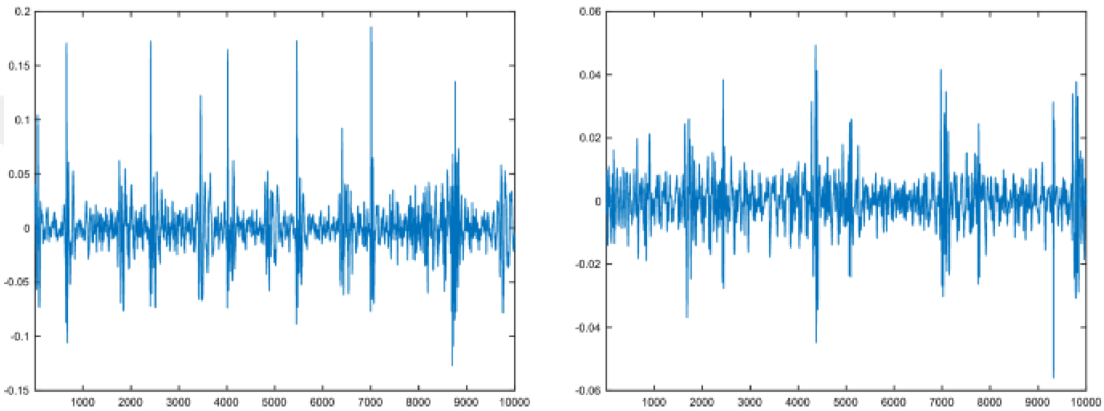
3. MEDICAL IMAGING AND SIGNALS

Medical Imaging Techniques (MITs) are medical procedures that do not involve the introduction of instruments into the body whatsoever. These techniques are used to see into the organ without opening the body physically. It's a device that aids in the diagnosis and treatment of a variety of medical issues. There are numerous medical imaging and signal approaches, each with its own set of dangers and benefits [36]. Thus, the objective of this research is to present a particular technique using deep learning. This technique is using heart sounds transformed into Electrocardiography (ECG) signals which are further transformed to scalogram images. What is a scalogram? It is defined as the actual number of a signal's ongoing discrete wavelet coefficient. These signals are sensitive to Interference. Thus, we transform signals into frequency domains which makes it more efficient to analyze even when noisy. This transformation process often makes signal analysis more complex, which means the existing simple classifiers could result in a poor performance [37].

Using scalogram images of the ECG signal as initial study data used in deep learning applications. This exhibits optimal results for the classification of morphological imagery. A signal is a function of one or more variables that transmits relevant data. When signals are received from living things. These signals are frequently referred to as biological signals since they carry information about that living system's condition and function. The pulse of patients, the voltage collected by electrodes positioned on the scalp, and hence the spatial information of X-ray absorption received from a CT scan are all examples of medical signals. In Machine learning applications we can use signal processing to automate measurement characteristics, reduce subjectivity and increase reliability of various signals. Another purpose of signal processing could be to filter out unwanted or noisy signal components that are either technical or physiological in nature, to make the examination of the signal part easier, for example, there have been several studies that have investigated heart rate variability, valve abnormality, identifying abnormalities and pathologies of the heart [11]. These are all various ways deep learning applies to medical signals.

Signals are either one-dimensional (i.e., they rely upon one variable (e.g., time), or multidimensional (e.g., space). It makes no difference if the signaling variables are temporal or some kind analytically [13]. Medical signals processing is often employed in monitoring real-time signals which regularly ends up in effective management of chronic diseases, detection and diagnosis of adverse diseases like heart attacks and strokes et al. Medical signals are especially useful within the critical care environment, where patient medical data must be analyzed in real-time to provide medical care [14] [15].

Object recognition (e.g., tissue placement), object segmentation (for example, lesion mapping), and classification are the most extensively used implementations of DL in diagnostic imaging and signal processing. (An example is Heart sounds classification using hybrid CNN architecture) [38]. Object identification, for example, has been extensively employed in computer-assisted diagnosis of radiography images to suggest probable cancers, as well as in liver - related CTs and X-rays. These applications also stretch to object segmentation because it's commonly used in computerized radiation therapy planning to demarcate cancers and organs as targets and amount of the drug respectively. Hence, whole radiology imaging and signals are usually supported by diagnostics classification [17]. As seen in figure 3.1. pictural representation of heart sound signal.



(a) Normal heart sound signal (b) Abnormal heart sound signal

Figure 3.1. Sample representation of heart sound signal; Normal **(a)** and Abnormal **(b)**.

3.1. ECG

Electrocardiogram (ECG) is basically some A particular type of graph that depicts cardiac electrical impulses from one moment to the next. The ECG is a period charting of the cardiac rhythm that is used in clinical assessment and therapy of patients as it can provide vital information. [39]. The ECG is a biological indicator that represents the electrical activity of the heart. It's known for tracking diseases with similar patterns in the way the heart beats, but it also often captures some beat regularities just like it does in reading signals from mental stress amongst others [40]. Electrodes placed on a patient's skin provide ECG readings, which indicate the heartbeats throughout time. To record the electrical potential coming from the heart in various angles and positions, the ECG leads are utilized to perform the task. The output often serves as indicators for condition through variance in the waveforms generated [41], [42].

ECG signals are simply characterized as records of the electrical facet of heart rhythm. As a result, it is now one of the most essential diagnostic techniques for cardiac disease. The ECG is an extremely valuable diagnostic tool, particularly in clinical situations. Due to human behavior and variances in the posture, size, and architecture of the heart, as well as other factors that may affect the output, ECG signals differ from person to person [43]. Studying ECG signals is a critical aspect of medical science because it contains a great deal of information [42], [44]. Therefore, prompt realization of abnormalities is crucial when it comes to monitoring and evaluation of patients in a medical setting [45].

The electrocardiogram (ECG) has been dynamically recognized as a popular diagnostic technology employed for the purpose of medical, clinical, and scientific research applications. The application of computers in processing of electrocardiograms (ECGs) has evolved greatly over the past 15 years, which increased rapidly [46]. This gave birth to opportunities for deep learning applications on ECG signals for classification purposes, a more dependable, automated, and low-cost tracking and diagnostics solution. The rise in adoption of deep learning (DL) techniques in the ever-growing area of systems used to assist doctors often known as “Computer-Aided Diagnosis Systems (CADS)”.

ECG signals processing has been an option for several decades in the implementation of supervised learning for medical applications, with several applications that include gathering information about the cardiac rhythm. This medical data is combined with understanding of both the heart's anatomy and physical findings indicators to be analyzed and used in diagnostic testing, prognosis, and medication. A multitude of programs and research work for DL methods with ECG have become greatly available [47].

3.1.1. Wavelet Transform (Scalogram)

To accurately identify heart sounds, determining the frequency of these extra noises in the rhythm is critical, distinguishing features among various diseases is also critical. Segmenting heart sound is an indispensable step in almost every understandable heart sound identification system [48]. This study uses the continuous wavelet transform (scalograms) to classify the abnormalities in the heart sounds.

A visual representation of continuous wavelet transforms (CWT) of a signal often known as Scalogram. This is a wavelet transform like a spectrogram that is created using a short time Fourier transform (STFT). Thus, researchers found concluded that CWT offers a more superior time and frequency resolution compared to the end results from STFT. Signals representing the frequencies that are present at varying times in the signal are called Scalogram. They also provide a visual representation that can be used to distinguish between the ECG signals for example heart sounds [48].

Generally, analyzing signals results in diagrams with continuous wavelet transform of varying frequency according to time. The method applied when converting to the time-frequency domain is very important when it comes to pattern recognition tasks. Thus, applying wavelet transform is most suitable for performing the transformation, because has proven to be effective in methods that relates to non-stationary signals such as ECG, EEG amongst other [49]. In this study, we transformed heart sound samples into images which can then be processed by a pretrained hybrid network. The images used are scalograms generated using wavelet transformation [50]. The intensity was represented on the y axis, and frequency was represented on the x plane, to make the scalogram images. To illustrate these generating visuals, we applied the Gouldian transformation algorithm. However, these varies from blue representing (low range) to green indicating (mid-range) to red showing (upper range) which gives the wavelet coefficient a visual representation using colors like a heat map. Furthermore, to ensure only the essential data on the images is fed into the proposed model, the margin markings were removed from all the generated scalograms as shown in figure 3.2 [51].

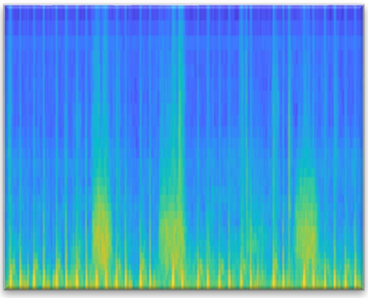


Figure 3.2. Sample scalogram image.

4. DEEP LEARNING

Deep learning techniques are based on algorithms (i.e., sets of mathematical procedures) that is implemented to describe the various relationships between variables. In this study, we will be explaining the conceptual overview and processes involved in DL techniques. This ML technique is a statistical tool for identifying patterns based on available data utilizing multiple layers of neural network models. When fed a suitably enough dataset, DL is a reliable means of boosting efficiency on a complicated system for describing mappings across endpoints [52], [53]. We will further discuss the procedures of developing (often referred to as training), evaluating (testing) and validating DL algorithms used for classification, feature extraction, prediction, and transfer learning. Our focus on this study as indicated in the title is mostly DL technologies implementation in medical domain. However, DL algorithms function in many ways depending on their nature, form, and type. There are some noteworthy attributes in the way these algorithms are developed and implemented. Even though the complexities of DL algorithms may appear abstruse, they often carry more than a low-key resemblance to ML algorithms and conventional statistical analysis. As seen in Figure 4.1., it illustrates the relationship between conventional statistical analysis (CSA), ML and DL in today's data science revolution.

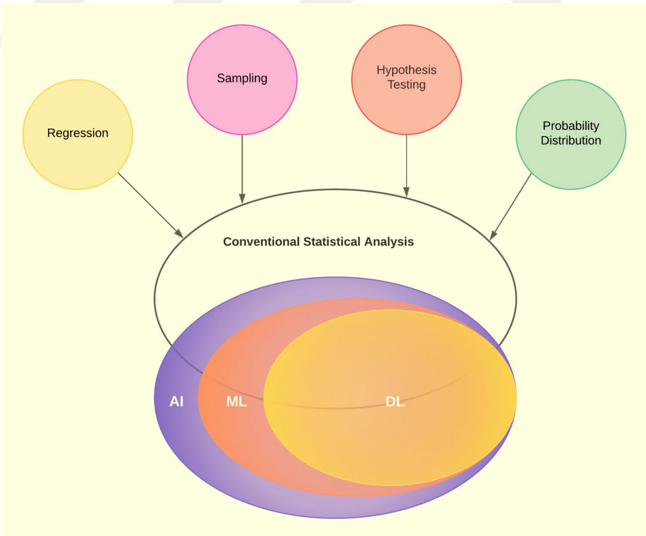


Figure 4.1. Relationship between statistical analysis, AI, ML and DL.

Conventional statistical models involve basic science of collecting, investigating, and presenting huge amounts of data to discover some hidden fundamental patterns and trends for making informed decisions. Statistics are often applied in practically almost everything we do daily, ranging from research, industrial and administrative application. A typical example is, medical practitioners and researchers keep patients healthy by using statistics to analyze data from the production of viral vaccines, which warrants consistency and reliability. The goal of conventional statistical methods is deduction and inference (i.e., to reach conclusions about populations or derive scientific knowledge from sample data of the population).

The conceptual framework of ML is a result of sets of computational proceeding which represents the connection between predictors or variables [54]. ML came to light from the study conducted on a method for automating some operations in the field of applied analysis by leveraging mathematical and mechanical actions. This came to reality with the answers to the following questions: Is it possible for a computer to go beyond what we can tell it to do and learn how to do a task by itself? Is it possible for such machine to learn set of rules by analyzing data? etc. Thus, ML algorithms are now a part of our day-to-day facets of life from what one can watch, to diagnose how healthy one is et al. For example, consider the application of ML algorithms in the detection of pneumonia in patients using medical data images (X-ray, CT scans etc.). These algorithms are deployed to predict whether traces of Pneumonia are present or not based on the consistent features seen in the studied dataset (known as training data). There are several other applications of ML algorithms, but this study focuses on its relationship to DL and medical science. ML develops and evaluates methods giving computers the ability to solve problems through learning from experiences. The main objective here is to create mathematically driven models that can be trained to produce essential outputs when given input data. These ML driven models are fed experiences in the form of training data, and then expected to be able to generalize their learned skill to deliver accurate prognosis on an entirely new data that is different from the training set by employing ML algorithms such as ANN, SVM et al. [55].

Furthermore, ML algorithms are categorized into two main categories. The first category being supervised learning. In this ML category, the aim is to predict some output variable that is linked with each input item. Thus, this needs to have training data with labeled objects to make the necessary predictions. The other type is, unsupervised learning category, where input data don't have any labels to go with the dataset. This kind of learning approach allows us to handle problems with insufficient or no knowledge about the final output [54].

In deep learning, machines learn essential representations and features with little to no direct human control, right from the raw data input, skipping the usual manual processes involved in ML and all the hectic steps requiring human control. The most known models in deep learning are many variants of neural networks, however, there are several others we'll discuss in detail [55]. DL algorithms necessitate two or more variable quantities validations of expertly tuned models [56]. DL technique has the potential to transform the medical science arena. Although, a considerable level of expertise in the field is essential to train such models to provide a desirable outcome. Some of the main characteristic of DL methods include its focus on feature learning (i.e., automatically learning representations of data) and feature extraction methods. These major differences between deep learning approaches and the commonly known machine learning methods makes the technique a more enhanced version of the latter. In DL methods, features extraction and carrying out a task is combined into one place. This enables both to improve during the same training process [57]. Thus, automating the entire feature extraction and classification process. Given in Figure 4.2., is a pictorial representation of a general deep neural network model.

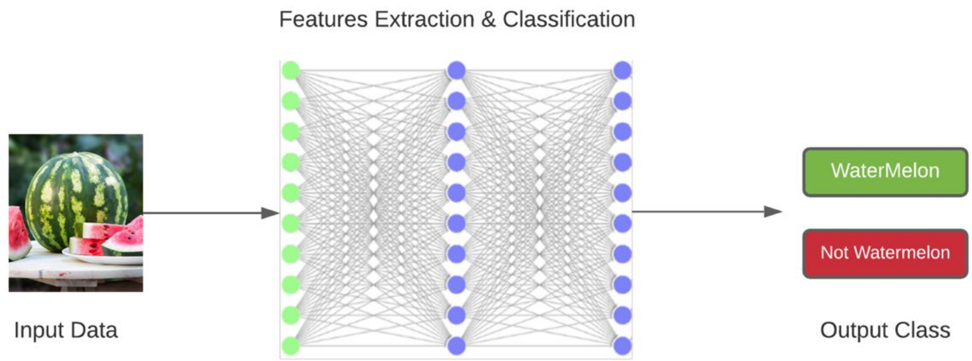


Figure 4.2. Example of a standard deep learning model.

All these technologies share something in common, which is to help in a better understanding of data used for classification, prediction, and analytics. However, each of these techniques takes a very different approach to achieving the desired goal. Understanding the difference between them will provide a clearer picture on their harmony. As a result, we see that the ideal approach is that these technologies are interrelated in a way. With all, attributes shared between statistical analysis, ML and DL techniques, may contribute to making the boundary between them seem vague or ill-defined. Thus, one way to present these set of approaches is to examine their main objectives and applications.

4.1. Neural Network (NN)

Neural networks are set of expressive networks. These are set of neurons used in machines, they are like the human brain neurons. These neuron-like networks are often employed in several branches of STEM. Neural networks constitute of parts connected in a group of basic operational pieces and components, that their purpose is similarly built to replicate the human neural system. The networks approach is kept in the link strengths between two or more units of the network, or weights, which are mostly gotten through conversion or learning representation. These kinds of networks are mostly used for the purpose of conducting statistical and data analysis/modelling, where they are seen and used as an alternate technique. These neural networks are often applied on classification and forecasting related problems. Typical examples of NN include image recognition, speech recognition, emotion recognition and NN applications in the medical domain for diagnosis and analysis et al. As Figure 4.3. has indicated, it illustrates simple neural network in which data is passed through sequence of layers before a prediction is computed. Advantages and disadvantages of neural networks are shown in Table 4.1.

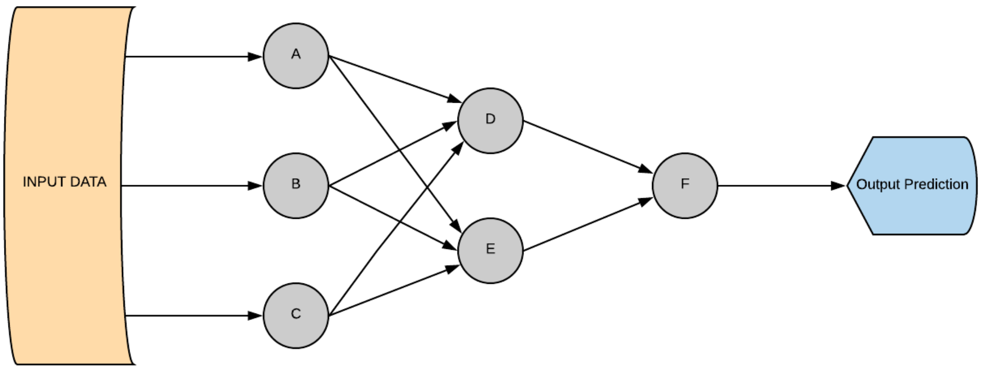


Figure 4.3. A sample neural network.

Table 4.1. Advantages and disadvantages of a neural network

Advantages	**Disadvantages**
They exhibit high speed. High accuracy and tolerance of noisy data.	It involves long training.
It is very much suitable for real world data.	Requires many parameters as topology.
The cost of maintainance and implementation in parallel hardware is good.	The process is difficult and requires high processing time if there are large neural networks.

4.1.1. Convolution Neural Network (CNN)

CNNs are highly regarded one of the widely adopted forms of Neural Networks, which basically are gotten from Artificial Neural Networks with several layers that are arranged in a fashion to take an input and give an output that links to the next layer in the network up until the last and final output classification layer. This kind of DNN is among the most powerful DL tools used in the field of Artificial Intelligence because of its ability to process huge sum of data. CNNs have been deployed in several practical applications such as, pattern recognition systems, digital image processing and analysis, voice, and emotion recognition systems et al. The results achieved from these applications were disruptive. This gave the technology a great reputation for yielding high performance in real life applications like healthcare systems, where it is used in medical data analytics and diagnostics among other things.

CNN is well known for its quality of using convolution layers to replace common matrix multiplication functions in at least one or multiple layers of the network to delicately show nonlinear diversified functions [58]. CNN multi-layer feature involves many connections, which gives the technology the capability to handle large amount of data. The structural representation of CNN is made up of various level of layers that include convolution blocks, pool blocks, activation layers, FC blocks, and some normalization functions.

Convolutional Neural Network possess a particular skill or behavior through practice and instructions called learning. In this deep learning technique, the learning process is often carried out using convoluted architectures. In CNN architectures, the input and output of every step in the network are basically in the form of numerical arrays known as feature maps, where all feature maps at the output point indicates specific feature(s) extracted via a given filter on the input of that layer. Moreover, these resulting features in higher phases are generally in a smaller form than the regular shape in a much lower stage, showcasing more practical features extracted [59].

There are several types of CNN pretrained networks with varying architectures. However, these architectures follow the same universal structure of applying convolutional layers to the input data to generate feature maps. Here are some of the commonly employed CNN architectures that are used for feature extraction in handling several deep learning tasks ranging from practical applications down to carrying out advanced tasks. These CNN architectures include AlexNet, GoogLeNet, LeNet, ResNet, VGG et al. As seen in figure 4.4., graphical representation of a basic convolutional neural network is made of different important layers.

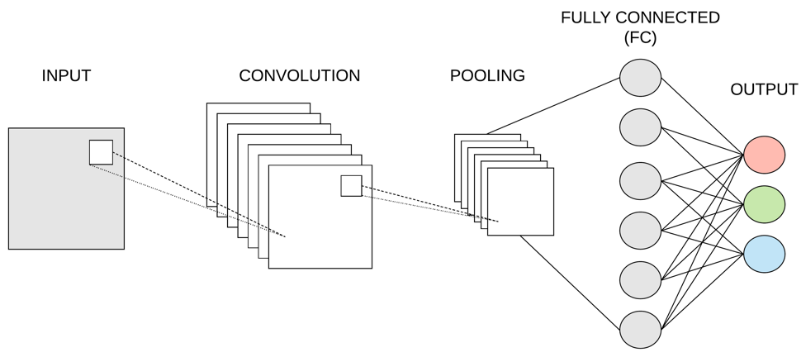


Figure 4.4. General representation of CNN architecture.

4.1.2. How do convolutional neural networks work?

These types of neural networks are easily differentiated from other types of neural networks based on their significant high performance on image, signal, or audio inputs. Convolutional neural networks consist of three major layers namely.

Convolutional layer

As the name implies, convolutional layer is the most critical component in a convolutional neural network. This layer is often the first part of the CNN and the core building block of any CNN. This layer checks and detects the range of attributes existing mostly in vector that has been fed directly. Convolution filtering, which really is a method of dragging a sort of filtering pane indicating similar features on the real input vector or image, is the core theory of the convnet. A cross products between both the features of the input image is then calculated. The features and filters are then seen as similar terms in this context. This layer houses most of the computation that is done in the process of training or testing a CNN. However, just like any other ML/DL based network, it requires components such as, input data, filter(s), and feature map(s). These networks also include a component called feature detector, often known as a kernel or a filter. How this component work is that it conducts a procedure called convolution, which is basically the movement across receptive fields of the input image, checking if the feature can be found in the input.

Moreover, the inputs received are processed and the convolution layer then computes the convolution of the input using the filter given. These filters must correspond with the target features we are trying to find and extract from the input. These two components make up the feature maps, and this then tells us where features with higher value are. The more corresponding input feature place in the input, the higher the value which indicates the features match the desired feature.

Convolution is a technique for preserving pixel relationships by extracting neural features which make comprise the pixels from tiny squares of training dataset. As seen in figure 4.5, this is a mathematical model. It takes two inputs which are the pixels and filter/kernel.

- Scale of input pixels $(h * w * d)$
- Filter size $(f_h * f_w * d)$
- Output matrix scale $(h - f_h + 1)\ x\ (w - f_w + 1) * 1$

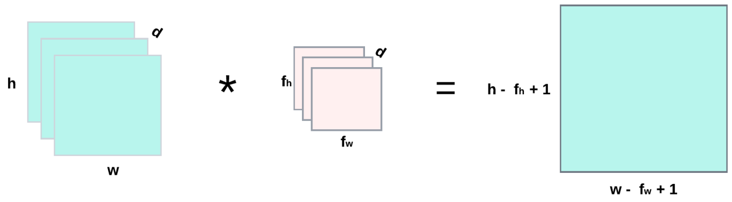


Figure 4.5. Filtering matrix result of given input array.

Assume a 5 by 5 bounding boxes with gray levels of 0 to 1 and a 3×3 filtration system vector, as seen in Figure 4.6.

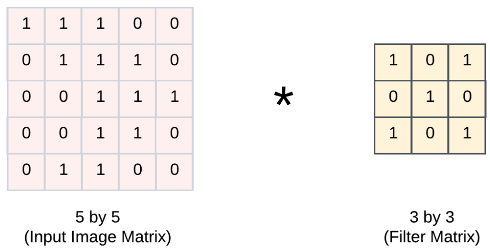


Figure 4.6. Input image matrix multiplied by the filter matrix

The next step in the process is the convolution of the 5 by 5 input image matrix multiplied with the 3 by 3 filter/kernel matrix known as the feature map which makes the output of convolved feature as shown in Figure 4.7 below

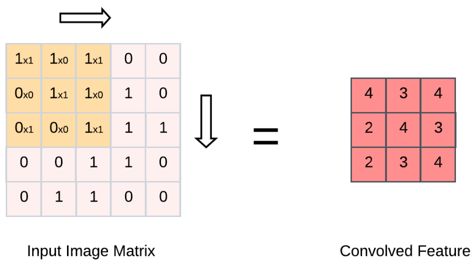


Figure 4.7. Convolved 3 by 3 output feature matrix

A CNN vary based on the way its layers are arranged and parameterized (i.e., these layers have hyperparameters that must first be defined with values). The convolutional layer in a CNN includes four (4) hyperparameters:

1. K, number of filters
2. F, size of filters (i.e., $F\ by\ F\ by\ D$ pixels define the single filter)
3. S, this represents the steps that drags the window corresponding filter across the input image (e.g., 2 is adjuting the filter frame two pixels every time it moves).
4. P adds an empty padding of P pixels in Zeros to the input data. By adding this contour, it makes the dimensions of the output fit the layer size. However, having many convolutional neurons arranged with $P = 0$ results in a input image, this results in loosing lots of information quickly, making the feature extraction process a lot more difficult.

Pooling layer

This layer of the convolutional neural networks is known for down sampling is a technique for reducing the number of variables (i.e., removing outliers in the input). Like the convolution block, in the pool block, a filter goes through the input gotten from the previous layers. However, the difference here is that there are no weights in the pooling layer. An aggregate function is used instead. is applied on the vectors by kernel within the receiving area, which then populates the list of output matrix from the pooling layer. The position of the polling layer in the CNN is frequently used together with two convolutions, in which it takes a few features maps and afterwards performs a pooling operation on them. What the pooling operation entails is basically reducing the number of image features extracted while keeping the important parts of these input data, which enhances the network's ultimate speed and reliability whilst avoiding overlearning or overfitting. There are two major types of pooling in CNN architectures:

A. Max pooling:

The max pooling technique is used more often in comparison to the average pooling. In this approach, the filter moves across the input in a similar way the filter moves across the input in the convolutional layer. However, here it selects the image pixel that has the maximum value which is then sent to the output array of the layer.

B. Average pooling:

In the average pooling approach, the filter moves across the input just as it does on the max pooling approach, but in this technique, it computes the average value within the receptive field which is then sent to the output array of the pooling layer.

Just as indicated earlier, a CNN is differentiated mostly by the arrangement of layers and their parameters. Therefore, convolution and pooling layers have hyperparameters that are defined to perform a task. There are two hyperparameters the pooling layer depends on:

1. F, the cell size on photos, which is further split into squares of dimension F by F pixel resolution.
2. S, this separates each cell by S pixel squares.

The biggest pixel from the smoothed feature space is used as the input for two - by - two filtered image in the pooling technique. Thus, average pooling can also be taken from the extracted feature from the convolution later. The summation pool, and that is the total over all bits in the feature map, is another option. Figure 4.8 below shows how using max pooling approach to subsample or each map's scale is reduced yet significant data is retained in the image.

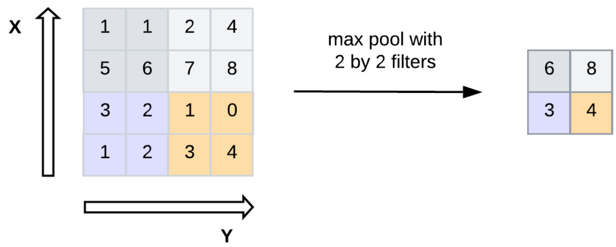


Figure 4.8. Max pooling approach.

Fully connected layers

These layers of the CNN aptly describe themselves. In a fully - connected layer, the image pixels of input data are tied directly to a node in the output neuron and the previous layer, as the name implies. The role this layer plays in a CNN is to perform the classification task based on the features extracted using different filters. Previous layers such as convolutional and pooling layers sometimes uses ReLU, “real non-linear” functions defined by $ReLU(x) = \max(0, x)$ (i.e., a correction layer that acts as an activation function that replaces all low model parameters to 0s). For the FC layers, the activation function used in classifying inputs accordingly is the SoftMax, resulting in a probability from 0 to 1.

This layer is always the last layer that makes up the final part of a neural network regardless of it being a convolutional based network. What this layer does is receive input vectors and produce new output vector after applying linear combinations and activation function to the accepted input vectors. Classification is done on the final FC layer: it returns vector of size N, (i.e., number of classes in the classification task). Each vector's element represents the likelihood that the input vector belongs to one of N classes. The FC layer finds the probability distributions to compute the probabilities which is the product of multiplying the data value by the weight, add them together, then use the activation function (e.g., SoftMax if $N > 2$). The idea of fully connected came to be because of input values connected with all output values. The FC layer determines the link between the locations of features retrieved in the training dataset and the outcome classification. In Figure 4.9 below, the feature map matrix gotten from the pooling layer is converted into vector (x1, x2, x3…, xN), which in the end is flattened as FC layer to classify the outputs accordingly.

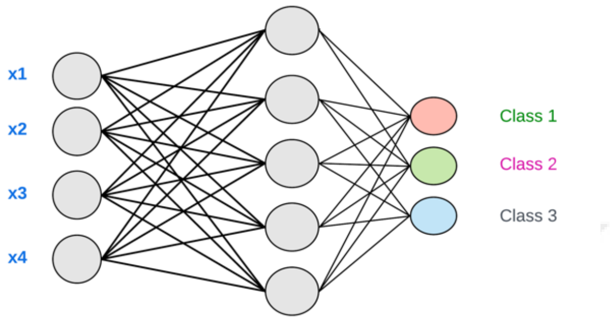


Figure 4.9. Illustration of fully connected layer.

4.1.3. Types of convolutional neural networks

AlexNet

A pre-trained network with a comprehensive feature encoding that can classify a wide range of photos from several applications. AlexNet CNN was designed/developed in the year 2012 by a machine learning researcher with the name Alex Krizhevsky in collaboration with other researchers, Ilya Sutskever and Geoffrey Hinton. These researchers had interests in the fields of AI, ML and DL. Upon development of the pretrained network, it competed in the ImageNet competition in 2012. The basic AlexNet structure is relatively like that of the LeNet-5, even though, this model is substantially bigger. The success of AlexNet began when it emerged first place winner in the 2012 ImageNet competition, this achievement helped in convincing the AI community to greatly adopt deep learning techniques for application purposes on computer vision tasks [60]. This network has eight weighted layers, the first five of which are convoluted and the last three of which are fully connected layers [61]. AlexNet input layer is fed with an input that has a size of (227 x 227 x 3). A typical AlexNet network structure is broken down as seen in Table 4.2.

Table 4.2. Model architecture of pre-trained AlexNet CNN.

Layer	Layer type	Layer Name	Layer Information
Input	'Image'	'input'	output size 227 by 227 by 3
1	'convolution'	'conv1'	11 by 11 kernel size, stride 4, 55 by 55 by 96 output, ReLU activation functions

Table 4.2. (Continued).

Layer	Layer type	Layer Name	Layer Information
	'Max pooling'	'pool1'	27 x 27 x 96 output size, 3 x 3 kernel size, stride 2, ReLU activation functions and padding [0 0 0 0]
2	'convolution'	'conv2'	27 x 27 x 256 output size, 5 x 5 kernel size, stride 1, ReLU activation functions and padding [2 2 2 2]
	'Max pooling'	'pool2'	13 x 13 x 256 output size, 3 x 3 kernel size, stride 2, ReLU activation functions and padding [0 0 0 0]
3	'convolution'	'conv3'	13 x 13 x 384 output size, 3 x 3 kernel size, stride 1, ReLU activation functions and padding [1 1 1 1]
4	'convolution'	'conv4'	13 x 13 x 384 output size, 3 x 3 kernel size, stride 1, ReLU activation functions and padding [1 1 1 1]
5	'convolution'	'conv5'	13 x 13 x 256 output size, 3 x 3 kernel size, stride 1, ReLU activation functions and padding [1 1 1 1]
	'Max pooling'	'pool5'	6 x 6 x 256 output size, 3 x 3 kernel size, stride 2, ReLU activation functions
6	'Fully-connected layer'	'fc'	9216 output size, ReLU activation functions
7	'Fully-connected layer'	'fc'	4096 output size, ReLU activation functions
8	'Fully-connected layer'	'fc'	4096 output size, ReLU activation functions
Output	'Classification layer'	'output'	SoftMax activation functions

GoogLeNet

GoogLeNet pre-trained CNN based network utilizing a 22-layer network and a picture feed dimension of 224 by 224 by 3. This network was also learned just on ImageNet database, which categorizes the input data into 1000 categories, including pictures of computers, mugs, crayons, and a variety of other items and animals. The GoogLeNet (Inception Network) is commonly used because it learned various attributes rendering techniques for a variety of images [62]. This network is known as Inception architecture because it is based on estimating the ideal local feature topology in a convnet while using dense layer components. GoogLeNet was chosen as a homage to Yann LeCuns who pioneered the development of LeNet-5 pretrained network [63]. The Inception network was introduced in the year 2014 by a team of researchers at Google. This network took first place in the 2014 classification and detection competition on ImageNet challenge. The Inception model is made up of a core unit called an "Inception cell". This unit is used to perform a series of convolutions at various scales and then collect the results in one location. In this model, 1 by 1 convolutions are used to reduce the input channel distance from top to bottom so that computations are saved. Hence, for every cell on the input, the network learns series of 1 by 1, 3 by 3, and 5 by 5 filters that are used in extracting features at distinct weights from the input parameters [62]. A typical GoogLeNet (Inception) architecture is represented as shown in Table 4.3.

Table 4.3. Model architecture of pre-trained GoogLeNet CNN.

Layer	Layer Type	Layer Information	Layer Depth
1	Input	224 x 224 x 3	1
2	Convolution	112 x 112 x 64 with patch size 7 x 7 and stride 2	1
3	Max Pool	56 x 56 x 64 with patch size 3 x 3 and stride 2	0
4	Convolution	56 x 56 x 192 with patch size 3 x 3 and stride 1	2
5	Max Pool	28 x 28 x 192 with patch size 3 x 3 and stride 2	0
6	Inception (3a)	28 x 28 x 256-layer output size	2
7	Inception (3b)	28 x 28 x 480-layer output size	2
8	Max Pool	14 x 14 x 480 with patch size 3 x 3 and stride 2	0
9	Inception (4a)	14 x 14 x 512-layer output size	2
10	Inception (4b)	14 x 14 x 512-layer output size	2
11	Inception (4c)	14 x 14 x 512-layer output size	2
12	Inception (4d)	14 x 14 x 528-layer output size	2
13	Inception (4e)	14 x 14 x 832-layer output size	2
14	Max Pool	7 x 7 x 832 with patch size 3 x 3 and stride 2	0
15	Inception (5a)	7 x 7 x 832-layer output size	2
16	Inception (5b)	7 x 7 x 1024-layer output size	2
17	Average Pool	1 x 1 x 1024 with patch size 7 x 7 and stride 1	0
18	FC	1 x 1 x 1024	0
19	Dropout	1 x 1 x 1024	0
20	Linear	1 x 1 x 1024	1
21	SoftMax	1 x 1 x 1024	0
22	Output		

LeNet

This part of the research paper explains the LeNet architecture in detail. The development of LeNet CNN model was pioneered by Yann Lecun's in the year 1998. The LeNet CNN is among the first few CNN models proposed by researchers in the field of AI. This convolutional neural network is made up of three (3) convolutional layers, two (2) pooling, one (1) FC layer, one (1) final classification layer, making in total seven (7) layers deep. Furthermore, activation functions are added between the NN layers to solve the problem of not being able to perform nonlinear operations. These activation functions basically serve as a trigger to the neuron via a function called mapping, just like in other forms of CNN method. It further improves the capacity of handling nonlinear issues that arises in the neural network. LeNet has an input size of 32 by 32 image. The general LeNet architecture is represented below, as seen in Figure 4.10. and as seen in Table 4.4., it's a basic representation of the model structure of LeNet-5 CNN architecture.

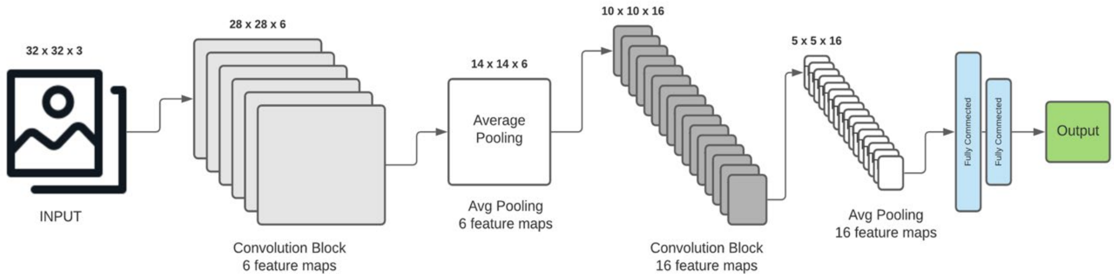


Figure 4.10. LeNet-5 architecture

Table 4.4. Model architecture of LeNet-5 CNN.

Layer	Layer Name	Layer Type	Output Size	Feature Map	Kernel Size/Stride
1	Input	Image	32 x 32	1	
2	'Conv1'	Convolution	28 x 28	6	5 x 5 / 1
3	'Avg Pooling'	Average Pooling	14 x 14	6	2 x 2 / 2
4	'Conv2'	Convolution	10 x 10	16	5 x 5 / 1
5	'Avg Pooling'	Average Pooling	5 x 5	16	2 x 2 / 2
6	'Conv3'	Convolution	1 x 1	120	5 x 5 / 1
7	FC		84		
8	FC	Output	10		

ResNet (Residual network)

Kaiming He and other researchers developed the RNN, a sort of fully convolutional network [64]. This idea gave the rise to the development of much deeper networks with several hundreds of layers unlike the typical tens of layers. The approach of the residual CNN based network is basically one that adds alternate routes as a link on every two layers, just like the architectural representation in a VGG network approach [65]. In the AI era, is believed that residual networks are easier to train because they successfully result in lower training and test errors in most of its applications. A basic example is team of researchers supervised the training of a network with 1001 layers using residual networks [66]. This process was not feasible to carry out before the emergence of Resnet, however with this technology it is practically possible to carry out several complex computer vision application tasks nowadays. Resnet happen to have a similar concept of alternate routes just like in LSTM methods. Albeit they don't use the concept of gates like the ones in LSTM recurrent networks. Thus, having a more polished shortcut route on Resnet helps in optimizing the function of the network to make them perform much better than any regular network [64], [67]. There are several Resnet with various added layers such as 18-layer, 34-layer and 101-layer residual networks et al. Nonetheless, the basic structure remains the same all through. As illustrated in Figure 4.11., we can see a general architecture of a residual network.

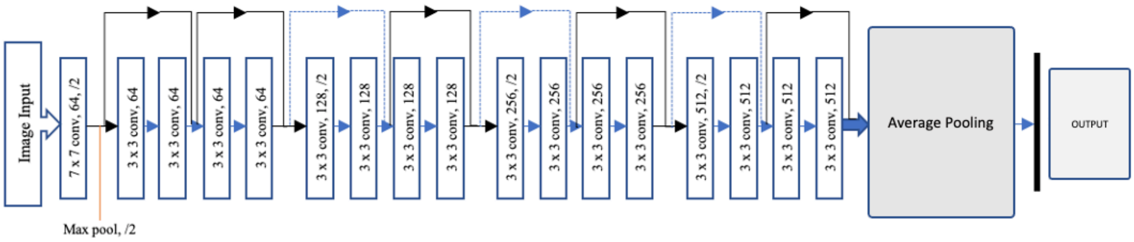


Figure 4.11. ResNet-18 Architecture.

VGG

VGG is short for (Visual Geometry Group). VGG is a form of convolutional neural network that came into existence in the year 2014. This CNN based network extends far back with a much deeper yet easy to understand type of the convolutional structures discussed in this section of the study. When the network first emerged, researchers and industry experts regarded it as the network with the most depth, however the narratives have changed in recent times, with the evolution in DL and its techniques giving more room for the creation of more robust networks with hundreds of layers. There are several types of VGG with different level of layers. However, they all share the same basic architecture which is made up of thirteen (13) convolutional layers, three (3) fully - connected layers, pooling layer, and thus the Linear activation function (like that of AlexNet) appended to these layers. VGG-16, for example, is made up of 16 layers, while VGG-19 is made up of 19 layers, and so on. This CNN-based VGG network takes in (224 x 224 x 3) pixel image as an input parameter. As shown in Figure 4.12., which represents a general architecture of VGG network.

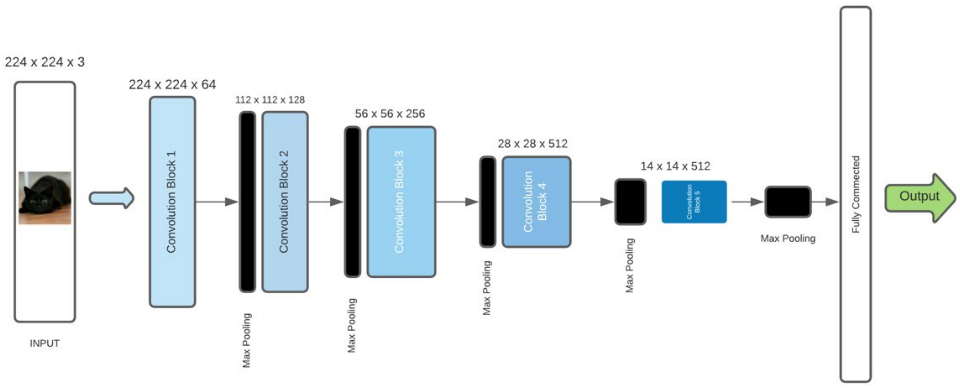


Figure 4.12. General architecture of VGG network.

4.1.4. Recurrent Neural Network (RNN)

We start off with a basic definition of the term recurrent. In simple words, "recurrent" basically mean occurring often or repeatedly in a sequence. In a Recurrent Neural Network (RNN) the pattern is essentially when the output at present time becomes the new input to the time event that follows. Here, the network focuses on the element of the cycle, where the RNN model examines the preceding elements of the input. To understand RNN better, it is a type of NN method under the umbrella of deep learning techniques, which processes sequences of elements, one at a time while keeping a memory often referred to as "state". This memory is of the previous element in the chain. RNN is basically a neural sequence method that sets to yield high performance on significant computer vision problems [68]. The AI research community have developed massive interest in research and development of RNN since 1990's. The basic building idea behind RNN is that they are made to study sequential or time-series events. Hence, making RNN a neural network with response links. Thus, RNNs come in many different forms.

Some practical applications of RNN include handling problems dealing with trajectories, control systems, adaptive robot behaviors language learning, chaotic systems, sequential auto association, filtering, and control of signal processing applications et al. Examples of RNN include Hopfield, bidirectional associative memory (BAM), echo state network (ESN), independently recurrent neural network (IndRNN), Long short-term memory (LSTM). This study only discusses LSTM RNN in detail. A basic RNN architecture is represented as seen in Figure 4.13.

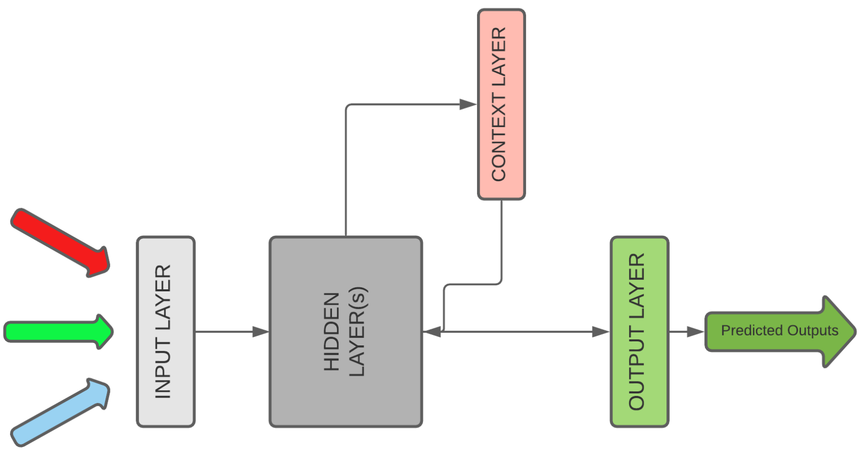


Figure 4.13. Comprehensive RNN Architecture.

Long Short-Term Memory (LSTM) network

LSTM is known for its ability to learn and memorize extended dependencies; it is a type of recurrent neural network which uses memory cell [69]. These RNN-based networks are also popular for their standard conduct of remembering information gone for long, which makes it highly capable of retaining information excessively. Like other types of recurrent networks, LSTM is a useful tool mostly deployed in time-series prediction problems due to its functionality of remembering preceding inputs. LSTM networks basically describes in detail the DL applications, what we do to handle using sequence and series data. A practical example of this network is using LSTM program to classify time series data.

This has a structural representation like a chain structure, where four (4) layers communicating conveys input and output data in a more distinctive fashion. The applications of for classification and regression, Recurrent networks go further than sequencing and time series analysis and predictions, they are often deployed for speech recognition, pattern recognition, music composition, pharmaceutical development, medical diagnostics et al. [67], [70]

A typical LSTM RNN-based network core component includes an LSTM structure and a sequence layer The sequence input layer is responsible for loading sequence and data over the period into the network, while the LSTM layer is responsible for learning the long-term relationships between time frames of the data sets. An LSTM network is typically set up with the sequence input neurons first, then the LSTM structure, and then the fully - connected layers, SoftMax layer, and classification output neurons (like final layers in AlexNet CNN), these last 3 layers are used in predicting the class labels just as they do in several other DL networks. The basic building idea of LSTM is to handle the disappearing gradient issue that is often stumbled upon during training a conventional RNN. LSTM networks are efficient and effective in handling classification, processing, and prediction of sequential data. Hence, these kinds of recurrent neural networks are deemed one of the most powerful if not the most vigorous classifier there is in the AI, DL community. There are four components that make up a general LSTM unit, namely, cell: This component serves the purpose of remembering values over random timeframes; input gate, output gate, forget gate: what they do is basically to regulate the flow of data and information to and from the cell.

The LSTM network contains some special units referred to as data cells, which are typically found inside the recurring hidden neurons of an RNN-based LSTM network. The transient state of the LSTM is stored in these memory blocks, which are made up of memory blocks with self-connecting links. In the LSTM design, these memory blocks incorporate the components discussed before, namely input, output, and forget gates [71]. The input gate, as the name implies, regulates the movement of input activation functions that go straight into the memory cell. Likewise, the output gate oversees controlling the flow of cell activations into the rest of the LSTM network. To address a drawback that prevents LSTM models from processing uninterrupted streams of data, which are not divided into subsequences of inputs, the forget gate is attached to the memory block [67], [71]. As shown in Figure 4.14., a diagram representing a simple RNN-based LSTM network architecture used in solving a classification problem.

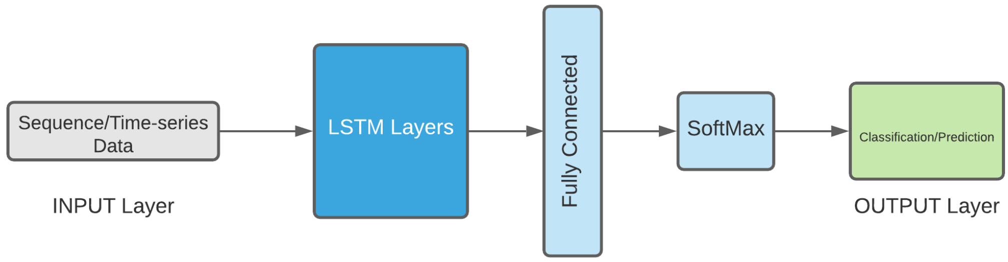


Figure 4.14. A simple LSTM network's structure

4.2. Feature Engineering

Deep learning for medicine has gained so much attention lately. Feature engineering techniques which include feature extraction and feature selection, often brings out hidden features in noisy input data. Firstly, the model performs feature processing, where data is fed to the model such as our proposed hybrid model. This process helps to identify and remove any from the dataset that may appear to be an outlier or redundant. It also helps in giving the raw data more structure and form for better processing. Additionally, feature extraction is a technique for reducing the scale of large datasets, which might impair learning process. In turn it speeds up the learning process[72]. This is a part of building reliable deep learning-based solutions for industries especially the medical field. This reduces the number of features that may appear to be useful in the learning process, which in turn reduces the computational power needed to train the proposed deep learning-based model. Some feature extraction techniques concerning time-frequency domain features is a wavelet transform (WT) like scalogram and spectrogram et al.

4.2.1. Feature Selection Algorithms

One of the characteristics of the feature engineering/processing techniques is that it enhances the effectiveness of a given DL model. This is a technique that has been applied in several studies related to mostly pattern recognition tasks amongst others. To improve the efficiency and accuracy of a machine learning system, feature selection is a method that selects a subset of attributes in each data set. This process allows the choosing of small subset of features that ideally describes the target outcome without fuss [73]. People often think the more features fed into a deep learning network, the more information there is for the machine to learn from. However, that is not the case, because some data features are noisy, invalid, and redundant. This is where feature selection algorithms like ReliefF algorithm come into play to create new feature vectors with valid and small amount of information for model training [74]. One of the main benefits of implementing selection algorithms in training a deep leaning-based model is the fact that it reduces model overfitting. As shown in Figure 4.15., feature selection is like playing dart. The minimal optimal method selects the best of the features and discard others.

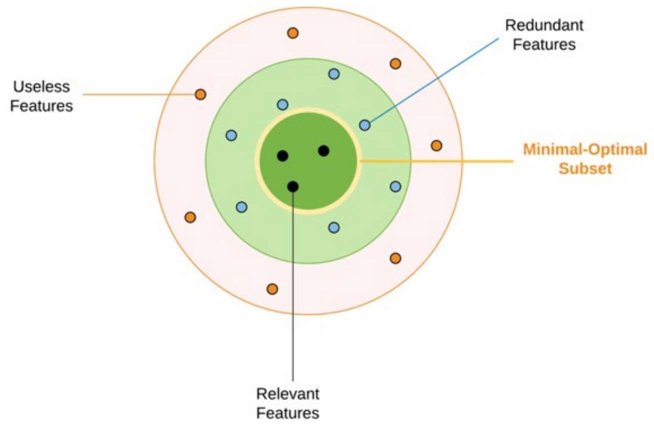


Figure 4.15. Illustration of feature selection.

Relief Selection Algorithm

This section of the study gives algorithmic and conceptual description of a type of feature selection algorithm formulated by Kira and Rendell [73]. A relief attribute selection algorithm is a set of approach for identifying and removing redundancy to minimize overfitting and increase performance of the proposed model. There have been significant improvements in model precision as well as a reduction in algorithm calculation time. The relief selection technique was employed in this study to identify the collection of variables that best differentiated irregular heart rhythms from regular heart sounds. In the case of a multiclass categorical variable, relief determines the scores of predictors. Predictors that offer different scores to neighbors in the same class are penalized, whereas predictors that give values to neighbors in distinct categories are rewarded [75]. The relief algorithms have primarily been utilized as a feature selection tool in feature engineering processes prior to model training. The main purpose of this algorithm is to get the number of quality of vectors according to their distinguishable values based on instances that are near to each other. This is seen in Figure 4.16., a more conceptual way as a pseudocode.

Required parameters

n *is the number of training instances*

a *is the number of features (i.e., attributes)*

m *is the number of random training instances out of n used to update* ***W***

Initialize all feature weights W[A] = 0;

for i = 1 to m do

Randomly select a target instance R_i;

find a nearest hit H instance and nearest miss M instance;

for A = 1 to a do

$W[A] = W[A] – diff(A, Ri, H)/m + diff(A, R_i, H)/m;$

end for

end for

return #Newly selected vector that estimates the quality of features.

Figure 4.16. Pseudocode for basic relief function.

Note: The difference between the values of Attribute for two instances is calculated using the function diff (Attribute; Instance1; Instance2).

There are two categories of feature selection algorithm namely, the filter and the wrapper methods. Researchers have determined that the wrapper method outperforms the filter method, but its application in the medical field is necessarily limited because of its high needs for computational power [76]. With regards to this study, a filter-based selection algorithm was applied. Furthermore, Relief algorithm is well-known for its ability to judge the quality of qualities based on how well their values differ from those of their immediate neighbors [77]. Relief algorithm employed in this study is a filtering featured selection method because the algorithm calculates for each feature, a proxy statistic that can be used to measure feature quality or relevance to the predicting endpoint value is calculated. These features are associated with statistics known as feature weights or feature scores, which ranges from -1 (worst instance) to +1 (best instance) [75]. The original Relief technique, however, had a constraint in that it could only be used to solve binary classification issues. Extended Relief algorithm called ReliefF was introduced to capture multi-class problems as well.

This extended version of the basic relief algorithm enables incomplete, multiclass, and noisy datasets, to use this function. ReliefF is a fast method that extracts attribute quality and strong relationships between them. Relief’s main notion is to estimate the best features by comparing them to similar situations [78].

Laplacian scores selection algorithm

Laplacian feature selection algorithm ranks feature in unsupervised learning using Laplacian scores. This method of selection has properties that make it reasonably impervious to outliers and noise [79]. This selection technique is hinged upon the concept of Locality Preserving Projection [80] and Laplacian Eigenmaps [81]. What this concept does is to evaluate the extracted features according to what is called the locality preserving power which is then used to evaluate the importance of the features extracted to select more accurate [79]. To get feature's locality preserving power, the Laplacian score is computed to match its LPP to determine good features with points nearby.

The whole procedure of the Laplacian score algorithm is as follows:

Input:

The value of score L_r as r^{th} feature.

f_{ri} shows the i^{th} case in r^{th} feature, $i = 1, \dots , m$.

Algorithm:

Step 1: Construct a nearest neighbor graph G with m nodes where the i^{th} node corresponds to X_i. We then put an edge between nodes i and j if X_i and X_j are near, (i.e., X_i is among k nearest neighbors of X_j or X_j is among k nearest neighbors of X_i.

Step 2: In this stage, where nodes I and j are connected, we choose the weights.

t is an appropriate constant, and S the nearest neighbors matrix gotten from wights in graph G in which the data space is represented.

Step 3: Computation for Laplacian score [81].

$$L_r = \frac{\widetilde{f_r}^T L \widetilde{f_r}}{\widetilde{f_r}^T D \widetilde{f_r}} \tag{4.1}$$

Output:

After calculating the Laplacian score for each feature and ranking them, we then get the important ranking of features to make the best selection possible in our extracted features.

Neighborhood Component Analysis (NCA)

The importance of feature selection is of considerable significance in deep learning, especially when it comes to dealing with high dimensional data. This method is known for maximizing the prediction accuracy in deep learning algorithms. It uses distance metrics to compute weights of the features selected [82]. NCA is a supervised feature selection technique that is developed using kNN classification algorithm [83]. Some of the benefits of this selection technique is its ability to generate positive weights for every feature. NCA learns a feature weighting vector and during the process of dimensionality reduction, this selection algorithm doesn't lose any information. This is a selection technique that provides information of significant features with their feature ranking details [83].

T as training samples, where x_i is a feature vector of dimension $d, y \in 1, \ldots$, selects x_j as its reference point as indicated in Equation (4.2).

$$p_{ij} = \begin{cases} \frac{k\left(D_w(x_i, x_j)\right)}{\sum_{k \neq i} k\left(D_w(x_i, x_j)\right)} & if\ i \neq j\ or\ 0, if\ i = j \end{cases} \tag{4.2}$$

Note: $k(.)$ is a kernel function.

Minimum Redundancy Maximum Relevance (MRMR) Selection Algorithm

Among all other available feature selection algorithms, this is a filter-based method that has often been used in relevant research work [84] that indicates its great popularity because of its high-level accuracy, although it may be computationally expensive to implement. This selection technique grows linearly in relation to the sample data, and it scales quadratically with the number of features [85]. Furthermore, because it is designed to determine an ideal set of features that are mutually and maximally distinct and can effectively represent the response variable, this selection process produces an ordered ranking of all the features. The MRMR selection process has several advantages, including minimizing the rate of feature set redundancy and successfully maximizing the relevance of a feature set to the response variable. The MRMR algorithm ranks features in the following order: [86]:

A. $\max_{x \in \Omega} V_x$ choose the feature with the greatest relevance. After that, the chosen feature is added to an empty set S.
B. Zero redundancy and the relevance features in S:
    a. Check for features with nonzero importance, if none, then skip to number 4.
    b. Else, choose features with maximum relevance, $\max_{x \in S^c, W_x = 0} V_x$. This is then added to S.
C. Repeat the process in step 2 for all features with redundancy not equal to zero.

D. Choose the feature with the highest MIQ value with nonzero relevance and redundancy in, then add it to the set S.

$$\max_{x \in S^c} MIQ_x = \max_{x \in S^c} \frac{I(x,y)}{\frac{1}{|S|}\sum_{z \in S} I(x,z)} \tag{4.3}$$

E. Repeat the process on step 4 till all features in S^c have no relevance.

F. The final step is to randomize the value of S.

Note: If the conditions mentioned in a step are not met, the algorithm may skip that step.

Furthermore, using the mutual information of features and the response, the MRMR feature selection algorithm additionally assesses redundancy and significance. The MRMR feature selection algorithm is also often used for ML/DL-based classification problems [84], [87]. In this study, the pseudocode of MRMR is presented in Figure 4.17. Summary of several feature selection algorithms are also seen in Table 4.5.

```
Input
    candidates                //This is the set of initial features
    numFeaturesWanted         //This is the number of selected features
Output
    selectedFeatures          //The set of selected features.
Initialization
    for feature fi in candidates do
        relevance = mutualInfo(fi, class);
        redundancy = 0.
        for feature fj in candidates do
            redundancy += mutualInfo(fi, fj);
        end for
        mrmrValues[fi] = relevance - redundancy;
    end for
    selectedFeatures = sort(mrmrValues).take(numFeaturesWanted);
```

Figure 4.17. Pseudocode for Original MRMR feature selection algorithm [85].

Table 4.5. Summary of feature selection algorithms.

Selection Algorithms	Supported Problem	Supported Data Type	Training Time
Relief	Classification and regression	Categorical and continuous features	Moderate
Laplacian	Unsupervised learning	Continuous features	Moderate
NCA	Classification	Continuous features	Moderate
MRMR	Classification	Categorical and continuous features	Fast

4.2.2. Classification Methods

There are a handful of classification algorithms available, including kNN, decision tree, MLP and SVM etc. Some of these algorithms are popular, due to their ability to make classification based on proximity, number of layer neurons amongst other factors[88].

In this study, the classification process is done when the features are extracted and selected using the CNN algorithm that results in a new feature vector. The neural network-based model is programmed to split its input into a set of categories. A classification neural network's output is usually after the application of SoftMax, with a few output neurons equal to the total number of classes received as input parameters. We replaced the basic pretrained deep learning classification algorithms from our pretrained model which were then replaced by a SVM classifier. Table 4.6., below shows us some examples of classification algorithms applied in deep learning projects.

Table 4.6. Example classification algorithms.

Classifier	Parameter & Values
SVM	Kernel (Gaussian or Linear) c
kNN	Euclidean distance (ED) k
Naïve Bayes	Probability/Minimum threshold
Decision tree	Number of leaves (Nodes)

Support Vector Machines (SVM) Classifier

Because it seeks to find the best subspace with a different dimension for separating distinct classes to the best of its ability, SVM is one of the most extensively used classification algorithms. Most importantly it often helps when performing multi-class classification, just like in this study. Furthermore, a great advantage of this classification method is that it is often free from being controlled by the size, and state or characteristic of possessing dimensions of the data used.

However, in this study, a multiclass SVM was fitted using the features collected from the training images. The following equations represent the SVM classifier function. For training data from the **x**^th^ and the **y**^th^ classes, (Lin, 2002).

$$\min_{w^{xy},b^{xy},\xi^{xy}} \quad \frac{1}{2}(w^{xy})^T w^{xy} + C \textstyle\sum_t \xi_t^{xy} \tag{4.4}$$

$$(w^{xy})^T \emptyset(v_t) + b^{xy} \geq -1 + \xi_t^{xy}, \ if \ y_t = x, \tag{4.5}$$

$$(w^{xy})^T \emptyset(v_t) + b^{xy} \leq -1 + \xi_t^{xy}, \ if \ y_t = y, \tag{4.6}$$

$$\xi_t^{xy} \geq 0$$

Therefore, using such function, if the notation $[(w^{xy})^T \emptyset(v_t) + b^{xy}]$ says (v) is in the x^{th} class, then the vote for x^{th} class is added by one. Else, the $\mathbf{y}^{\mathbf{th}}$ class is increased by one and then we predict that (v) is in the class with the largest vote.

k-Nearest Neighbor

The implementation of the k-Nearest Neighbor (kNN) classification algorithm is discussed in this section. The input variables (also known as predictors, characteristics, or attributes) and output variables form the basis of this classification process (response). As classification algorithm, k-nearest neighbors (kNN) are simple yet interesting algorithm, these are supervised machine learning algorithm that are often employed in solving both classification and regression problems in ML. The implementation of kNN algorithm is easy although it has some setbacks of being slow when applied on huge data. The approach is based on an instance-based lazy learning algorithm for predicting the dominant class of a testing observation among the k most similar examples (nearest neighbors) in each challenge. The object is classified using the k-Nearest Neighbor (kNN) classifier based on the majority vote of its neighbors, with the object being allocated to the most common class among its k-Nearest Neighbors [89]. As a result, it assigns a point to the class among its k closest neighbors (where k is an integer value) [90]. Using kNN, we can easily determine the class of a particular dataset using its nearest neighbors. This is seen in Figure 4.18.

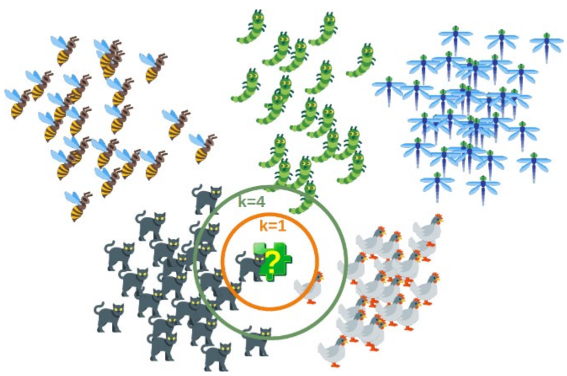


Figure 4.18. Sample classification using k nearest neighbors.

The steps below explain how kNN algorithm work:

1. The initial stage is to decide on the number of neighbors K.
2. Secondly, we then determine the distance between K neighbors.
3. In accordance with Euclidean distance computed, take the K nearest neighbor.
4. In each category, count the number of data points of the K neighbors.
5. Assign the new data points to the categories with the maximum number of neighbors.
6. Finally, the model is ready for use.

Furthermore, nearest instances are computed by calculating the distance used to classify the unknown instance(s) represented by the feature vectors extracted using our pre-trained CNN, where the (kNN) classifier computes the distances between the various points in the training data. The CNN is considered a point in the feature space. As a result, the Euclidean distance is chosen as the distance metric in this study while employing the k-Nearest Neighbor classifier to classify our dataset. Equation (4.7) is a representation of the Euclidean distance formula:

$$dist(A,B) = \sqrt{\frac{\sum_{i=1}^{m}(x_i - y_i)^2}{m}} \tag{4.7}$$

Assume that A and B are extracted feature vectors.,

$$A = (x_1,\ x_2, \dots, x_m) \text{ and } B = (y_1,\ y_2, \dots, y_m)$$

Where, the dimensionality of the feature space is m,

∴ to calculate the distance between point and points, we use the normalized Euclidean metric as described in Equation (4.7) above.

Naive Bayes Classifier

The Bayesian network classifier is a probabilistic classifier that uses Bayes' theorem to classify data. It computes the posterior probability which is often used for classification of data [86][91]. This classification algorithm is one of the oldest classifiers in ML tasks, it applies density estimation when performing the classification task. It uses the Bayes rule to obtain variables (assuming features) that are independent of each other given the class. The independence assumption of NB is an advantage, albeit it has proven to have good classification performance for many data sets that are real, especially for documents-based data. The classification algorithm takes these steps [92]:

1. Predictor's densities in each class are estimated.
2. Using the Bayes rule, calculate the posterior probability. i.e., for all $k = 1, \dots, k$ values, where Y is the random variable relative to the class index of an observation, $X_1, \dots, X_p$

are the random predictors of an observation, and k an indication of the class index, where $\pi(Y = k)$ is the prior probability.

$$\hat{P}\left(Y = k \middle| X_1, \dots, X_p\right) = \frac{\pi(Y=k)\prod_{j=1}^{p} P(X_j|Y=k)}{\sum_{k=1}^{K} \pi(Y=k)\prod_{j=1}^{p} P(X_j|Y=k)} \tag{4.8}$$

3. Classify an observation according to the class yielding maximum posterior probability.

Decision Tree Classifier

Decision tree learning is one of the most widely used practical approaches for inductive inference. As a robust approach of approximating discrete-valued target functions, the learned function is shown in the form of a decision tree. To increase human readability, trees can alternatively be represented as conditional statements, such as a collection of if-then rules. This algorithm is a well-known inductive inference process that is used for a variety of applications, including classification and medical diagnosis. From the root node to the leaf nodes, the algorithm organizes the hierarchy of decision nodes down to the prediction of classes. In the classification algorithm based on decision trees, ending with the classification, the instances are arranged by sorting through the tree starting from root to some leaf nodes or vice versa. As shown in Figure 4.19., a typical representation of a learned decision tree. This decision tree classifies Thursday mornings according to their suitability for playing tennis.

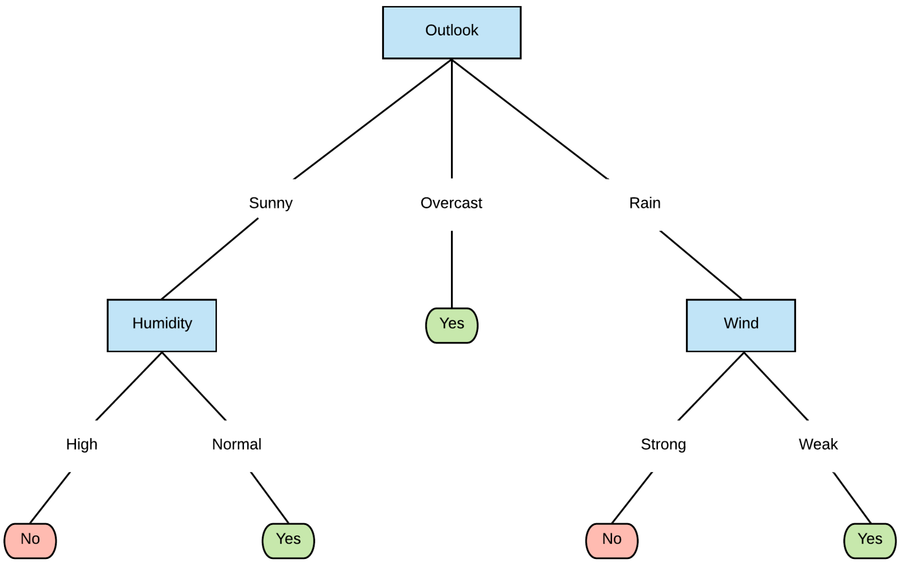


Figure 4.19. Sample decision tree.

We see the basic concepts in a decision tree with the implementation algorithm (Pseudo-code) behind decision tree classifier as shown in Figure 4.19.

- Training examples are the Examples.
- The Target attribute is the one whose value the tree is supposed to anticipate.
- The Attributes are a collection of additional qualities that the learnt decision tree can examine.
- As a result, a decision tree is returned that appropriately classifies the target Examples.

- Create the tree's root node.
- If the statements used as examples are all positive, then return the root of a single-node tree, with label = +
- If the examples offered are negative, with label = -, return the single-node tree root.
- If the attributes associated with example are empty, a node tree with the label "most common value of Target attribute" is returned.
- Else begin
  - Attributes that best classifies the given examples are A
  - Root's decision attribute = A
  - For every potential value, v_i, of A,
    - Create a new tree branch that corresponds to the test $A = v_i$
    - Let Examples $_{vi}$, be the value-added subset of Examples v_i for A
    - If Examples $_{vi}$ is blank
      - Add a leaf node beneath the newly created branch = Target attribute's most common value in Examples
      - If not, add the subtree underneath the new branch.
- **End**
- **Return *Root***

Note: This algorithm continues until the training instances are perfectly classified by the tree, or until all characteristics have been exhausted.

5. APPLICATION RESULTS AND DISCUSSION

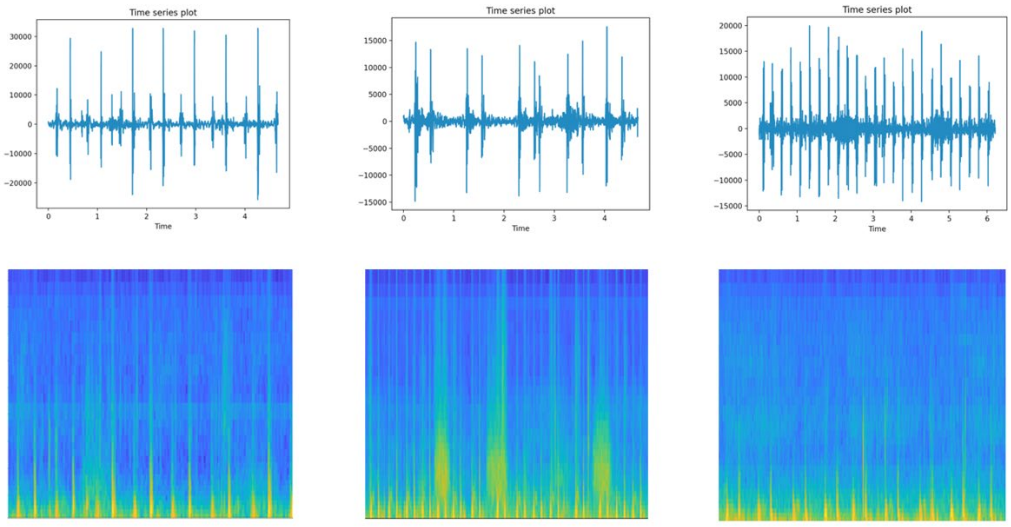
Cardiac abnormalities are regarded as one of the life most threatening abnormalities today. As reports indicates by WHO, heart diseases are the main reason behind most of the death cases today. It is estimated that cardiovascular diseases are behind the death of about 17.9 million people. This accounts for roughly 31% of all deaths worldwide, with heart attacks and strokes accounting for 85% of all deaths. [93]. Most heartbeat abnormalities are reflected in heart sounds. These sounds are listened to and interpreted by doctors using a stethoscope. This method is simple and inexpensive. In addition, by recording heart sounds, sounds can be converted into phonocardiogram graphics. This chart is then used by doctors for disease diagnosis. The accuracy of the diagnosis in these techniques are heavily reliant on the physician's knowledge. Recent years have seen an increase in systems that can automatically classify heart sounds and assist doctors have been developed. The necessity for assistive methods for diagnosing heart conditions has increased. DL algorithm such as CNN, are most ordinarily applied to investigate visual imagery. In the study, an intelligent computer aided diagnosis system is developed. The algorithms for this application are CNN-based. Several studies suggest that diagnostic tools such as medical signals contain salient information about heart abnormalities.

5.1. Data

This study contains a dataset containing heart sounds converted to scalogram images using wavelet transforms. The dataset is categorized into three categories: extrasystole, murmur and normal. The dataset consists of 346 total scalograms. The data used was gotten from the Peter J Bentley online repository. This information was collected from two sources: the public via the iStethoscope Pro iPhone app and a clinic trial in hospitals utilizing the Digi Scope, a digital stethoscope. The lengths of the audio heart sounds vary between 1 and 10s. The dataset is obtained under clinical and non-clinical settings. This data set contains several variations in the hardware used for recoding the heart sound. The quality and type of patient also created a challenging environment for evaluation.

However, the study's main goal is to use heartbeat patterns to automatically detect aberrant heart sounds including murmurs and extrasystole disorders. [38].

The portable network graphics (PNG) format is used for all scalogram images. Figure 5.1. shows sample raw audio signal and scalogram from the three categories applied in this study.



(A) (B) (C)

Figure 5.1. Raw sound signal and scalogram images (A) Extrasystole, (B) Murmur & (C) Normal.

5.2. The Proposed Solution

The proposed solution in this paper is essentially a Hybrid CNN model based on the Resnet50 and Resnet101 designs. As seen in Figure 5.2., this model is offered for the classification of heart sounds. As an input, it uses the scalogram images obtained from the heart sounds. Each image's feature vector is extracted from the Fc1000 layers, and the feature vectors are concatenated to create a new vector output. The relief feature selection technique is applied on the generated feature vectors to make selections from new samples. The features that maximize the classification's success are determined. Finally, a support vector machine classifier is used to classify the feature vector. The Hybrid CNN model created in this work is shown in Figure 5.2.

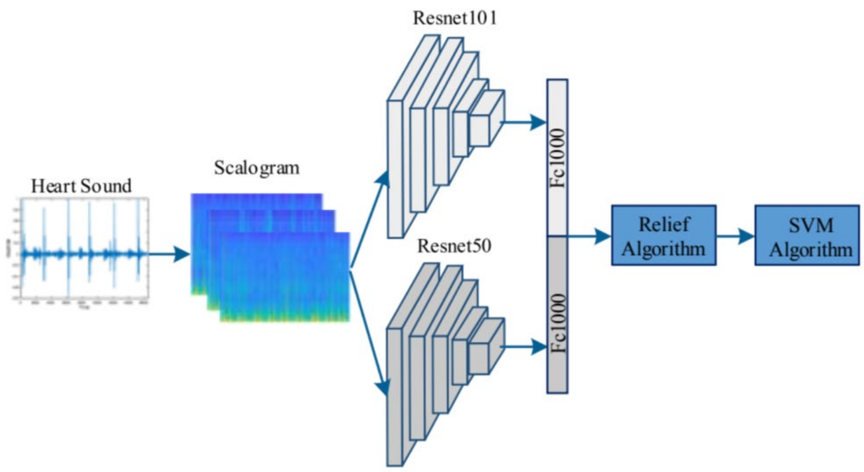


Figure 5.2. The proposed method.

5.2.1. Processes applied in the proposed model

The following process is applied for the Hybrid CNN method proposed for classifying heart sounds in a total of 346 wav format.

Step 1: Convert the wav file format to time series. Figure 3 shows an example heart sound signal.

Step 2: Time-frequency representations of the signals are generated. These representations are called scalograms. The absolute value of the continuous wavelet transform (CWT) is called a scalogram, which is the coefficients of a signal.

Step 3: Scalogram images are the inputs of Resnet50 and Resnet101 models.

Step 4: The new feature vector is created by combining the feature vectors obtained from the fc1000 layer of both models.

Step 5: Select feature vectors using relief algorithm and classify with Support Vector Machine (SVM).

5.3. Evaluation of Performance using metrics

The accuracy, precision, sensitivity (recall), specificity, and F1 score of the proposed method is evaluated. We'll use the following terms to better describe these evaluation metrics: TP (i.e., properly predicted positive values indicating that the value of the real class is yes, as well as the value of the predicted class) TN (properly predicted negative values, indicating that the value of the real class is no, and the value of the predicted class is no as well) When the real class is no but the projected class is yes, it is called FP. When the actual class is yes but the projected class is no, the result is FN. As a result, the proposed approaches' performance is assessed using the metrics given in this paper.

1. Accuracy is a measurement metric used to determine better models at mapping out patterns and relationships between variables in a dataset, which is often influenced by the input, training, or test data. The number of right predictions divided by the total number of forecasts equals accuracy, which is then multiplied by 100 to yield a percentage. Accuracy is calculated using Equation (5.1) below.
2. Precision in ML/DL means, how accurate the model is in classifying data sample as positive. In other words, the precision determines how effective and reliable the model is in terms of classifying samples as positive. Precision is calculated using Equation (5.2).
3. Sensitivity often referred to as Recall is the measure that tells the ability of the detection model for positive samples. In general, the better the model finds positive samples, the higher the recall score. It is calculated using Equation (5.3).

4. Equation (5.4) indicates how to mathematically calculate the specificity score. This is a measure that indicates the proportion of the actual number of samples classified as true negative. Specificity measure and sensitivity are also used in plotting the ROC curve and AUC curve which further helps to determine the performance of ML/DL-based models.
5. The harmonic Mean value between two important measures in evaluating a DL model is called the F1 Score. This indicates how the model is and how well does the model classify instances (i.e., does the model miss significant number of instances). Research indicates the greater the F1 Score on a model, the better the overall performance score of that model is. F1 Score is mathematically calculated using Equation (5.5).
6. The MCC is a metric for determining the relationship between actual classes and expected labels. This is mathematically calculated as seen in Equation (5.6).

It is shown in the results that the developed model could assist doctors in diagnosing heart disease and making clinical decisions effectively. A summary of the evaluation metrics is shown below.

Accuracy

$$\frac{TP+TN}{TP+FP+FN+TN} \tag{5.1}$$

Precision

$$\frac{TP}{TP+FP} \tag{5.2}$$

Sensitivity (Recall)

$$\frac{TP}{TP+FN} \tag{5.3}$$

Specificity

$$\frac{TN}{TN+FP} \tag{5.4}$$

F1 Score

$$2 * \frac{(Sensitivity*Precision)}{(Sensitivity+Precision)} \tag{5.5}$$

Matthews Correlation Coefficient (MCC)

$$\frac{(TP*TN-FP*FN)}{\sqrt{((\mathrm{TP+FP})*(\mathrm{TN+FN})*(\mathrm{FP+TN})*(\mathrm{TP+FN}))}} \tag{5.6}$$

5.4. Results

The data set was randomly generated for training, validation, and testing, respectively, 70%, 10%, and 20%. 1x1000 size feature vectors were obtained from FC1000 layer of Resnet50 and Resnet101 models used in the model given in Figure 5.2. The new feature vectors with 1x2000 dimensions were combined and given to the Relief feature selection algorithm. Then, the properties that maximize the classification accuracy are determined and classified. Figure 5.4 confusion matrices and Table 5.1., indicates the value of the obtained classification parameters. The total number of features for which the highest level of accuracy was achieved with the proposed method was determined to be 315. As seen in Figure 5.3., the proposed method's overall accuracy was determined to be 92.75%.

Table 5.1. Performance parameter values obtained.

Methods	Accuracy	Precision	Sensitivity (Recall)	Specificity	F1 Score	Matthews Correlation Coefficient (MCC)
ReliefF + SVM	92.75%	92.0%	92.3%	96.0%	92.1%	96.8%
Resnet50	86.9%	80.2%	81.8%	91.9%	81.0%	88.8%
Resnet101	84.0%	80.0%	83.4%	91.0%	81.7%	91.7%
Densenet201	69.6%	66.0%	67.7%	30.0%	64.3%	27.8%
Alexnet	75.36%	61.0%	63.0%	77.8%	59.0%	52.6%
Mobilenetv2	79.71%	69.7%	69.0%	50.0%	68.7%	37.0%

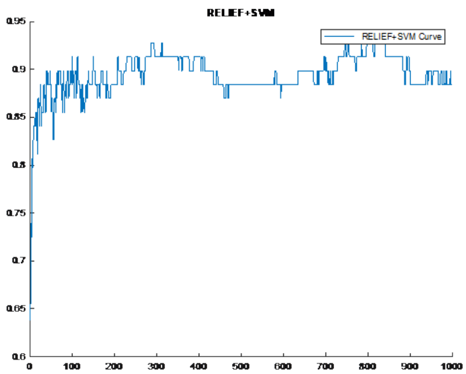


Figure 5.3. Accuracy variation of the proposed method according to the features used.

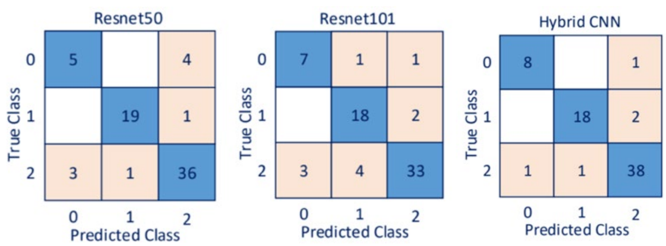


Figure 5.4. Confusion matrices

6. Conclusion, limitation and Suggestion for Future Work

6.1. Conclusion

As discussed in the study, deep learning models are a simple or complex composition of multiple hidden layer neural network trained to perform a task with less human input in the process of training, unlike a conventional machine learning model. This study will impact medical science field with the knowledge of practical applications of deep learning techniques for clinical and research purposes. This research work discusses the application of deep learning on medical images and medical signals for the diagnosis of heart abnormalities. Many discussed types of ML and DL techniques such as deep neural networks employed in algorithms. These DL techniques have accomplished astonishing achievements across medical science domain. Ever since the emergence of DL, the number of researchers has increased at an exponential rate alongside the number of industry professionals that became interested in exploring the various application of DL in computer-aided diagnosis. Thus, making DL a dominant approach in tackling variety of complicated medical conditions such as heart abnormalities, neurological disorders, brain tumors etc. DL is not a panacea in any of its applications, however, it is currently being studied and researched upon thoroughly, to improve its potential in medical science.

In this study, we discussed some practical applications of DL in heart disease diagnosis and the performance of these algorithms were measured using a general guideline on evaluating such machine learning techniques. Moreover, the study and application results evaluated shows how effective DL techniques are enhancing the research area of computer-aided cardiovascular disease diagnosis. It shows how various methods are applied to several image and signal data showcasing feature extraction using CNN models, selection algorithms and classifiers for effective training and prediction of heart disease affected patients. With recent events in the surge of cardiovascular mortality rate, this could serve as an alternative means to diagnose CVD patients. The application discussed has so much potential to highlight the key role DL techniques play in the medical field. Although, there are several high-impact publications in top-journals that discussed and reviewed the idea of deep learning in heart disease diagnosis. Hence, adoption of DL in clinical application comes with its own challenges and setbacks although it is highly beneficial.

Thus, this research work utilized CNN for collecting readable features from the time-frequency representations of the signals which are generated from heart audio sounds. The heartbeat sounds used in this study were a mix of recording gotten from a mobile phone app and a digital stethoscope. The hybrid method outperformed other pre-trained CNN models overserved in this research with an accuracy level of 93%. The experiments reveal that our state-of-the-art hybrid method could be deployed to assist doctors in diagnosing heart conditions at an early stage while aiding clinicians in taking decisions efficiently and effectively. However, the size of datasets used in this study is necessarily limited in number of samples.

6.2. Limitation

The most notable limitation of the computer-aided diagnosis system for heart disease includes the limited number of heart sounds dataset available for public use. Thus, creating a gap that calls for further investigation on a much larger dataset of denoised heartbeat sounds of different categories of cardiovascular diseases or heart conditions.

6.3. Suggestion for Future Work

The following suggestions can be taken into consideration to further encourage a more detailed, elaborate, and rich research on the subject matter. These suggestions to build on the future research work include:

1. Publications should further be examined and studied to keep innovating and enhancing DL applications in the medical field.
2. The attention DL is gaining from the medical community can also be leveraged to evolve the computational mindset among medical and technology researchers and professionals.
3. Further research on heart sounds using time-frequency analysis should be conducted to have a more reliable assessment on the performance of the developed model.
4. There’s also a need to implement such applications with other forms of networks such as (Bi-LSTM, RNN, fully custom model), with larger size of datasets, and

REFERENCES

[1] S. Bianchini, M. Müller, and P. Pelletier, "Deep Learning in Science *."

[2] R. E. Twogood and F. G. Sommer, "Digital Image Processing*," 1982.

[3] I. Antonopoulos *et al.*, "Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review," *Renewable and Sustainable Energy Reviews*, vol. 130. Elsevier Ltd, Sep. 01, 2020. doi: 10.1016/j.rser.2020.109899.

[4] I. Arel, D. Rose, and T. Karnowski, "Deep machine learning-A new frontier in artificial intelligence research," *IEEE Computational Intelligence Magazine*, vol. 5, no. 4, pp. 13–18, Nov. 2010, doi: 10.1109/MCI.2010.938364.

[5] I. Arel, D. Rose, and T. Karnowski, "Deep machine learning-A new frontier in artificial intelligence research," *IEEE Computational Intelligence Magazine*, vol. 5, no. 4, pp. 13–18, Nov. 2010, doi: 10.1109/MCI.2010.938364.

[6] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift fur Medizinische Physik*, vol. 29, no. 2. Elsevier GmbH, pp. 102–127, May 01, 2019. doi: 10.1016/j.zemedi.2018.11.002.

[7] C. Hodges *et al.*, *Handbook of Deep Learning Applications*, vol. 136, no. August. 2019. [Online]. Available: http://link.springer.com/10.1007/978-3-030-11479-4

[8] C. Shen, D. Nguyen, Z. Zhou, S. B. Jiang, B. Dong, and X. Jia, "An introduction to deep learning in medical physics: advantages, potential, and challenges," *Physics in Medicine & Biology*, vol. 65, no. 5, p. 05TR01, Mar. 2020, doi: 10.1088/1361-6560/ab6f51.

[9] V. Kaul, S. Enslin, and S. A. Gross, "History of artificial intelligence in medicine," *Gastrointestinal Endoscopy*, vol. 92, no. 4, pp. 807–812, Oct. 2020, doi: 10.1016/J.GIE.2020.06.040.

[10] S. Kaptoge *et al.*, "World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions," *The Lancet Global Health*, vol. 7, no. 10, pp. e1332–e1345, Oct. 2019, doi: 10.1016/S2214-109X(19)30318-3.

[11] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep Learning Applications in Medical Image Analysis," *IEEE Access*, vol. 6, pp. 9375–9379, Dec. 2017, doi: 10.1109/ACCESS.2017.2788044.

[12] B. Ozaydin, E. S. Berner, and J. J. Cimino, "Appropriate use of machine learning in healthcare," *Intelligence-Based Medicine*, vol. 5, p. 100041, 2021, doi: 10.1016/j.ibmed.2021.100041.

[13] Y. Zeinali and S. T. A. Niaki, "Heart sound classification using signal processing and machine learning algorithms," *Machine Learning with Applications*, vol. 7, p. 100206, Mar. 2022, doi: 10.1016/J.MLWA.2021.100206.

[14] W. Chen, Q. Sun, X. Chen, G. Xie, H. Wu, and C. Xu, "Deep learning methods for heart sounds classification: A systematic review," *Entropy*, vol. 23, no. 6. MDPI AG, Jun. 01, 2021. doi: 10.3390/e23060667.

[15] A. Castro, T. T. v Vinhoza, S. S. Mattos, and M. T. Coimbra, "Heart Sound Segmentation of Pediatric Auscultations Using Wavelet Analysis."

[16] L. Huiying', L. Sakari2, and H. Iiro', "A Heart Sound Segmentation Algorithm Using Wavelet Decomposition and Reconstruction."

[17] E. R. Mehmet Bilal, "Heart sounds classification using convolutional neural network with 1D-local binary pattern and 1D-local ternary pattern features," *Applied Acoustics*, vol. 180, p. 108152, Sep. 2021, doi: 10.1016/J.APACOUST.2021.108152.

[18] M. A. Alonso-Arévalo, A. Cruz-Gutiérrez, R. F. Ibarra-Hernández, E. García-Canseco, and R. Conte-Galván, "Robust heart sound segmentation based on spectral change detection and genetic algorithms," *Biomedical Signal Processing and Control*, vol. 63, p. 102208, Jan. 2021, doi: 10.1016/J.BSPC.2020.102208.

[19] A. K. Gárate-Escamila, A. Hajjam El Hassani, and E. Andrès, “Classification models for heart disease prediction using feature selection and PCA,” *Informatics in Medicine Unlocked*, vol. 19, p. 100330, Jan. 2020, doi: 10.1016/J.IMU.2020.100330.

[20] M. Deng, T. Meng, J. Cao, S. Wang, J. Zhang, and H. Fan, “Heart sound classification based on improved MFCC features and convolutional recurrent neural networks,” *Neural Networks*, vol. 130, pp. 22–32, 2020, doi: 10.1016/j.neunet.2020.06.015.

[21] M. Alkhodari and L. Fraiwan, “Convolutional and recurrent neural networks for the detection of valvular heart diseases in phonocardiogram recordings,” *Computer Methods and Programs in Biomedicine*, vol. 200, Mar. 2021, doi: 10.1016/j.cmpb.2021.105940.

[22] N. Baghel, M. Kishore Dutta, and R. Burget, “Automatic diagnosis of multiple cardiac diseases from PCG signals using convolutional neural network,” *Computer Methods and Programs in Biomedicine*, vol. 197, p. 105750, 2020, doi: 10.1016/j.cmpb.2020.105750.

[23] W. Zhang, J. Han, and S. Deng, “Abnormal heart sound detection using temporal quasi-periodic features and long short-term memory without segmentation,” *Biomedical Signal Processing and Control*, vol. 53, p. 101560, Aug. 2019, doi: 10.1016/J.BSPC.2019.101560.

[24] M. Hamidi, H. Ghassemian, and M. Imani, “Classification of heart sound signal using curve fitting and fractal dimension,” *Biomedical Signal Processing and Control*, vol. 39, pp. 351–359, Jan. 2018, doi: 10.1016/J.BSPC.2017.08.002.

[25] B. Bozkurt, I. Germanakis, and Y. Stylianou, “A study of time-frequency features for CNN-based automatic heart sound classification for pathology detection,” *Computers in Biology and Medicine*, vol. 100, pp. 132–143, Sep. 2018, doi: 10.1016/J.COMPBIOMED.2018.06.026.

[26] J. Rubin, R. Abreu, A. Ganguli, S. Nelaturi, and K. Sricharan, “Recognizing Abnormal Heart Sounds Using Deep Learning.”

[27] W. Zhang, J. Han, and S. Deng, “Heart sound classification based on scaled spectrogram and partial least squares regression,” *Biomedical Signal Processing and Control*, vol. 32, pp. 20–28, Feb. 2017, doi: 10.1016/J.BSPC.2016.10.004.

[28] M. M. A. Rahhal, Y. Bazi, H. Alhichri, N. Alajlan, F. Melgani, and R. R. Yager, “Deep learning approach for active classification of electrocardiogram signals,” *Information Sciences*, vol. 345, pp. 340–354, Jun. 2016, doi: 10.1016/J.INS.2016.01.082.

[29] M. M. Azmy, “Classification of normal and abnormal heart sounds using new mother wavelet and support vector machines,” *2015 4th International Conference on Electrical Engineering, ICEE 2015*, Feb. 2016, doi: 10.1109/INTEE.2015.7416684.

[30] S. W. Deng and J. Q. Han, “Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps,” *Future Generation Computer Systems*, vol. 60, pp. 13–21, Jul. 2016, doi: 10.1016/J.FUTURE.2016.01.010.

[31] S. K. Randhawa and M. Singh, “Classification of Heart Sound Signals Using Multi-modal Features,” in *Procedia Computer Science*, 2015, vol. 58, pp. 165–171. doi: 10.1016/j.procs.2015.08.045.

[32] V. N. Varghees and K. I. Ramachandran, “A novel heart sound activity detection framework for automated heart sound analysis,” *Biomedical Signal Processing and Control*, vol. 13, no. 1, pp. 174–188, Sep. 2014, doi: 10.1016/J.BSPC.2014.05.002.

[33] F. Safara, S. Doraisamy, A. Azman, A. Jantan, and S. Ranga, “Wavelet Packet Entropy for Heart Murmurs Classification,” *Advances in Bioinformatics*, vol. 2012, 2012, doi: 10.1155/2012/327269.

[34] S. Yuenyong, A. Nishihara, W. Kongprawechnon, and K. Tungpimolrut, “A framework for automatic heart sound analysis without segmentation,” 2011. doi: 10.1186/1475-925X-10-13.

[35] C. Ahlstrom *et al.*, “Feature extraction for systolic heart murmur classification,” *Annals of Biomedical Engineering*, vol. 34, no. 11, pp. 1666–1677, Nov. 2006, doi: 10.1007/S10439-006-9187-4.

[36] H. Kasban, M. A. M. El-Bendary, and D. H. Salama, “A Comparative Study of Medical Imaging Techniques,” 2015. [Online]. Available: https://www.researchgate.net/publication/274641835

[37] Y.-H. Byeon, S.-B. Pan, and K.-C. Kwak, “Intelligent Deep Models Based on Scalograms of Electrocardiogram Signals for Biometrics,” 2019, doi: 10.3390/s19040935.

[38] M. M. Abubakar and T. Tuncer, “Heart Sounds Classification Using Hybrid CNN Architecture,” Aug. 2021. doi: 10.52460/issc.2021.023.

[39] A. L. Goldberger, Z. D. Goldberger, and A. Shvilkin, “Goldberger’s Clinical Electrocardiography A Simplified Approach Ninth Edition.”

[40] Z. Ebrahimi, M. Loni, M. Daneshtalab, and A. Gharehbaghi, “A review on deep learning methods for ECG arrhythmia classification,” *Expert Systems with Applications: X*, vol. 7, p. 100033, Sep. 2020, doi: 10.1016/J.ESWAX.2020.100033.

[41] Q. Yao, R. Wang, X. Fan, J. Liu, and Y. Li, “Multi-class Arrhythmia detection from 12-lead varied-length ECG using Attention-based Time-Incremental Convolutional Neural Network,” *Information Fusion*, vol. 53, pp. 174–182, Jan. 2020, doi: 10.1016/J.INFFUS.2019.06.024.

[42] F. Murat, O. Yildirim, M. Talo, U. B. Baloglu, Y. Demir, and U. R. Acharya, “Application of deep learning techniques for heartbeats detection using ECG signals-analysis and review,” *Computers in Biology and Medicine*, vol. 120, p. 103726, May 2020, doi: 10.1016/J.COMPBIOMED.2020.103726.

[43] B. P. Simon and C. Eswaran, “An ECG Classifier Designed Using Modified Decision Based Neural Networks,” *Computers and Biomedical Research*, vol. 30, no. 4, pp. 257–272, Aug. 1997, doi: 10.1006/CBMR.1997.1446.

[44] Ö. Yildirim, “A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification,” *Computers in Biology and Medicine*, vol. 96, pp. 189–202, May 2018, doi: 10.1016/J.COMPBIOMED.2018.03.016.

[45] Z. Ebrahimi, M. Loni, M. Daneshtalab, and A. Gharehbaghi, “A review on deep learning methods for ECG arrhythmia classification,” *Expert Systems with Applications: X*, vol. 7, p. 100033, Sep. 2020, doi: 10.1016/J.ESWAX.2020.100033.

[46] J. L. Willems, P. Arnaud, J. H. van Bemmel, R. Degani\, P. W. Macfarlane, and C. Zywietz, “Common standards for quantitative electrocardiography: goals and main results. CSE Working Party.”

[47] H. H. M. Draisma *et al.*, “LEADS: An interactive research-oriented ECG/VCG analysis system,” in *Computers in Cardiology*, 2005, vol. 32, pp. 515–518. doi: 10.1109/CIC.2005.1588151.

[48] A. Meintjes, A. Lowe, and M. E. Legget, *Fundamental Heart Sound Classification Using the Continuous Wavelet Transform and Convolutional Neural Networks*. 2018. doi: 10.0/Linux-x86_64.

[49] A. Narin, “Detection of Focal and Non-focal Epileptic Seizure Using Continuous Wavelet Transform-Based Scalogram Images and Pre-trained Deep Neural Networks,” *IRBM*, Nov. 2020, doi: 10.1016/J.IRBM.2020.11.002.

[50] S. C. Olhede and A. T. Walden, “Generalized Morse wavelets,” *IEEE Transactions on Signal Processing*, vol. 50, no. 11, pp. 2661–2670, Nov. 2002, doi: 10.1109/TSP.2002.804066.

[51] Z. Ren, N. Cummins, V. Pandit, J. Han, K. Qian, and B. Schuller, “Learning image-based representations for heart sound classification,” in *ACM International Conference Proceeding Series*, Apr. 2018, vol. 2018-April, pp. 143–147. doi: 10.1145/3194658.3194671.

[52] G. Marcus *et al.*, “Deep Learning: A Critical Appraisal.” [Online]. Available: http://www.nytimes.com/2012/11/24/science/scientists-see-advances-in-deep-learning-a-part-of-artificial-

[53] S. Bianchini, M. Müller, and P. Pelletier, “Deep Learning in Science *.”

[54] J. A. M. Sidey-Gibbons and C. J. Sidey-Gibbons, “Machine learning in medicine: a practical introduction”, doi: 10.1186/s12874-019-0681-4.

[55] A. S. Lundervold and A. Lundervold, “An overview of deep learning in medical imaging focusing on MRI,” *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 102–127, May 2019, doi: 10.1016/J.ZEMEDI.2018.11.002.

[56] L. Faes *et al.*, “Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study,” *Articles Lancet Digital Health*, vol. 1, pp. 232–274, 2019, doi: 10.17632/rscbjbr9sj.3.

[57] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553. Nature Publishing Group, pp. 436–444, May 27, 2015. doi: 10.1038/nature14539.

[58] I. Namatēvs, “Deep Convolutional Neural Networks: Structure, Feature Extraction and Training,” *Information Technology and Management Science*, vol. 20, no. 1, Jan. 2018, doi: 10.1515/itms-2017-0007.

[59] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998, doi: 10.1109/5.726791.

[60] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks.” [Online]. Available: http://code.google.com/p/cuda-convnet/

[61] A. A. Almisreb, N. Jamil, and N. M. Din, “Utilizing AlexNet Deep Transfer Learning for Ear Recognition,” in *Proceedings - 2018 4th International Conference on Information Retrieval and Knowledge Management: Diving into Data Sciences, CAMP 2018*, Sep. 2018, pp. 8–12. doi: 10.1109/INFRKM.2018.8464769.

[62] C. Szegedy *et al.*, “Going deeper with convolutions.”

[63] V. Kaul, S. Enslin, and S. A. Gross, “History of artificial intelligence in medicine,” *Gastrointestinal Endoscopy*, vol. 92, no. 4, pp. 807–812, Oct. 2020, doi: 10.1016/J.GIE.2020.06.040.

[64] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Dec. 2016, vol. 2016-December, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[65] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” 2015. [Online]. Available: http://www.robots.ox.ac.uk/

[66] S. Targ, D. Almeida, and K. L. Enlitic, “Workshop track-ICLR 2016 Resnet in Resnet: Generalizing Residual Architectures.”

[67] H. H. Sak, A. Senior, and F. Beaufays Google, “Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition.”

[68] T. Mikolov and G. Zweig, “Context dependent recurrent neural network language model,” in *2012 IEEE Workshop on Spoken Language Technology, SLT 2012 - Proceedings*, 2012, pp. 234–239. doi: 10.1109/SLT.2012.6424228.

[69] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[70] X. Li and X. Wu, “Constructing Long Short-Term Memory Based Deep Recurrent Neural Networks for Large Vocabulary Speech Recognition,” 2015.

[71] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, “Learning Precise Timing with LSTM Recurrent Networks,” 2002. [Online]. Available: www.idsia.ch

[72] D. K. Soother *et al.*, “The Importance of Feature Processing in Deep-Learning-Based Condition Monitoring of Motors,” 2021, doi: 10.1155/2021/9927151.

[73] “AAAI92-020”.

[74] D. Koller, M. Sahami, and G. Building, “Toward Optimal Feature Selection.”

[75] R. J. Urbanowicz, M. Meeker, W. la Cava, R. S. Olson, and J. H. Moore, “Relief-based feature selection: Introduction and review,” *Journal of Biomedical Informatics*, vol. 85. Academic Press Inc., pp. 189–203, Sep. 01, 2018. doi: 10.1016/j.jbi.2018.07.014.

[76] V. Nikulin and G. J. Mclachlan, “On a General Method for Matrix Factorisation Applied to Supervised Classification.” [Online]. Available: http://www.broad.mit.edu/cgi-bin/cancer/publications/

[77] H. Liu, H. Motoda, and L. Yu, “A selective sampling approach to active feature selection,” *Artificial Intelligence*, vol. 159, no. 1–2, pp. 49–74, Nov. 2004, doi: 10.1016/J.ARTINT.2004.05.009.

[78] Z. Pang, D. Zhu, D. Chen, L. Li, and Y. Shao, “A Computer-Aided Diagnosis System for Dynamic Contrast-Enhanced MR Images Based on Level Set Segmentation and ReliefF Feature Selection,” 2014, doi: 10.1155/2014/450531.

[79] M. Belkin and P. Niyogi, “Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering.”

[80] X. He and P. Niyogi, “Locality Preserving Projections.”

[81] X. He, D. Cai, and P. Niyogi, “Laplacian Score for Feature Selection.”

[82] W. Yang, K. Wang, and W. Zuo, “Neighborhood component feature selection for high-dimensional data,” *Journal of Computers*, vol. 7, no. 1, pp. 162–168, 2012, doi: 10.4304/jcp.7.1.161-168.

[83] M. Djerioui, Y. Brik, M. Ladjal, and B. Attallah, “Neighborhood component analysis and support vector machines for heart disease prediction,” *Ingenierie des Systemes d’Information*, vol. 24, no. 6, pp. 591–595, 2019, doi: 10.18280/isi.240605.

[84] V. Bolón-Canedo, A. Alonso-Betanzos, and D. Rego-Fernández, “Scalability analysis of mRMR for microarray data Scalable machine learning algorithms: Beyond classification and regression View project SMARTEES: Social Innovation Modelling Approaches to Realizing Transition to Energy Efficiency and Sustainability View project Scalability Analysis of mRMR for Microarray Data.” [Online]. Available: https://www.researchgate.net/publication/286370006

[85] S. Ramírez-Gallego *et al.*, “Fast-mRMR: Fast Minimum Redundancy Maximum Relevance Algorithm for High-Dimensional Big Data”, doi: 10.1002/int.21833.

[86] C. Ding and H. Peng, “Minimum Redundancy Feature Selection from Microarray Gene Expression Data.” [Online]. Available: http://www.nersc.gov/~cding/MRMR/.

[87] H. Peng, F. Long, and C. Ding, “Feature selection based on mutual information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005, doi: 10.1109/TPAMI.2005.159.

[88] Institute of Electrical and Electronics Engineers, *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC) : Banff Center, Banff, Canada, October 5-8, 2017.*

[89] R. A. Minhas, A. Javed, A. Irtaza, M. T. Mahmood, Y. B. Joo, and A. J. Pk, “Shot Classification of Field Sports Videos Using AlexNet Convolutional Neural Network”, doi: 10.3390/app9030483.

[90] L.-Y. Hu, M.-W. Huang, S.-W. Ke, and C.-F. Tsai, “The distance function effect on k-nearest neighbor classification for medical datasets,” 2016, doi: 10.1186/s40064-016-2941-7.

[91] L. Fan, K. L. Poh, and P. Zhou, “A sequential feature extraction approach for naïve bayes classification of microarray data,” *Expert Systems with Applications*, vol. 36, no. 6, pp. 9919–9923, Aug. 2009, doi: 10.1016/J.ESWA.2009.01.075.

[92] T. M. (Tom M. Mitchell, *Machine Learning*.

[93] S. Kaptoge *et al.*, “World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions,” *The Lancet Global Health*, vol. 7, no. 10, pp. e1332–e1345, Oct. 2019, doi: 10.1016/S2214-109X(19)30318-3.

APPENDICES

APPENDIX 1: A BRIEF HISTORY OF COMPUTING & AI

In the world today, computers play an interesting role, impacting lives and everyday life affairs. It doesn't just do that, but it is impacting everything ranging from communication, transportation, agriculture, and medical sciences etc.

These electronic devices (computers) are often seen as mind-blowing modern invention involving electronic and technological components. However, computing predates the use of devices that works with electricity. It is known that the famous computing device such as the abacus was among the first of its kind to have ever existed but that doesn't make it the first computer. Analog computing has been in existence several millennia back because history shows that several devices from the Roman and Greek times have been used as early computing technologies. Among the popularly known computing devices were the Antikythera mechanism which was complex. Subsequently as technology grew and advanced in that time, computing devices such as the complex mechanical devices like the castle clock by Al-Jazari, and slide rule by William. These are all examples of early mechanical inventions that served as analog computers. However, there are several other examples of computing devices that dates ages and centuries back.

The history of AI goes way back to 1956, where two young mathematicians, John McCarthy and Marvin Minsky met with the famous Claude Shannon who was well known for inventing information theory and the well-known designer of IBMs early days computers, Nathaniel Rochester developed a study to proceed to simulate intelligence that machines can have their own human matched intelligence and further learn from it. The objective was to make machines use languages, solve problems that are reserved for human minds while improving themselves at it.

In the first decades AI research had so many breakthroughs such as, Alan Robinson's logical reasoning algorithm, and the program developed by Arthur Samuel used to play a checker game. The algorithm thought itself to beat its creator. Despite the advancements and successes in the field, there was some setbacks in the first phases of development of AI which happened in the late 1960s where efforts in making intelligent machines learn on their own failed to meet the expectation put on them.

The AI research community learned its lessons, became smarter and better with the application of more mathematical theories and functions. Starting 2011, deep learning technologies started some grate advances in speech recognition, object recognition tasks etc. In 2016 thru 2018, AI powered Deep Minds - AlphaGo defeated a former world, Champion Lee Sedol. With this development, AI became so popular with huge potential economic and social gains in the world.

APPENDIX 2: HISTORY OF AI IN MEDICINE

AI has had so many advancements in the past years, that became a huge part of the field evolving into subdomains such as machine learning, deep learning, and computer vision. Birth of this advanced level of intelligence has greatly impacted technology growth. Some of the areas that AI has evolved into include:

i. ML: The identification of patterns and analysis where machines improve themselves with readily available data.

ii. DL: This is a subset of AI that is made of multi-layer-based networks with a structure of the normal neurons inside the human brain. This neural structure enables the machine to learn and make decisions all by itself making tasks less demanding of human input.

iii. CV: Also, a subfield of AI that enables a computer or machine to gain information and develop an understanding from patterns and series of images, sequences, or videos.

AI advanced into ML techniques which gave birth to DL, which basically is a collection of algorithms that creates what is known as artificial neural network (ANN). The evolution process is seen in Figure A2.1., and some modern computing technologies can also be seen in Table A2.1.

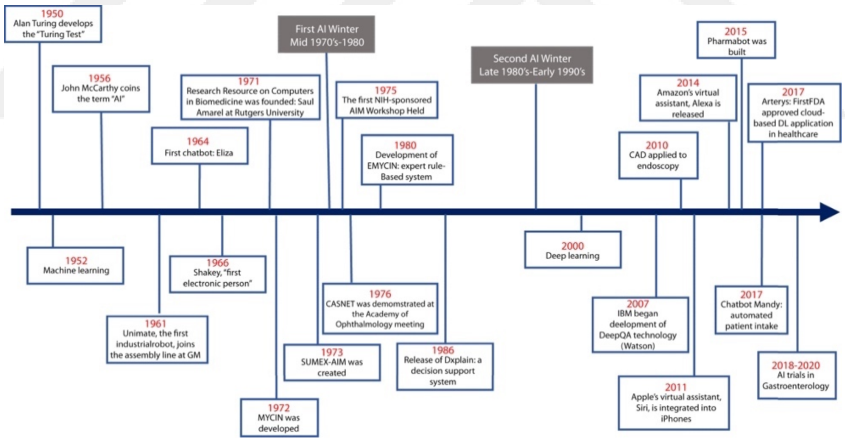


Figure A2.1. Representation of development in AI, ML & DL in medical sciences [94].

Table A2.1. Modern computing-based cardiovascular approved technologies [95]

S/N	Compliance No.	Owner	Product	Details	Approval Date
1	K140933	Alive Cor	Monitor System	System that checks for any cardiovascular related disease from ECG signals for vital signs of abnormal heartbeats.	15th August 2014
2	K160016	Steth IO	Mobile stethoscope App	Detects normal and abnormal heart conditions in real-time based on AI technologies.	15th July 2016

Table A2.1. (Continued). [95]

S/N	Compliance No.	Owner	Product	Details	Approval Date
3	K172311	Biotricity	Bio Flux	AI-based system that allows doctors to observe and monitor patients from home.	15th December 2017
4	K171056	Excel Medical Electronics	WAVE Clinical System	AI medical system developed to keep patients' data for monitoring and prediction.	4th January 2018
5	DEN180044	Apple	Mobile App	The company rolled out an app on their wearable device to detect heart conditions from the ECG signals read.	11th September 2018
6	K182344	Bio fourmis	Rhythm data app	The AI application developed to give a detailed analysis of heart patterns to detect abnormalities in ECG signals.	7th March 2019
7	K192004	Eko	Eko data system	System that assists doctors to give a proper evaluation of patient's heart conditions etc.	15th January 2020

CURRICULUM VITAE

Mohammed Mansur ABUBAKAR

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

ACADEMIC ACTIVITIES

Paper:

M. M. Abubakar and T. Tuncer, “Heart Sounds Classification Using Hybrid CNN Architecture,” Aug. 2021. doi: 10.52460/issc.2021.023.