Journal of Integrated
SCIENCE & TECHNOLOGY

A systematic review on Machine learning and Neural Network based models for disease prediction

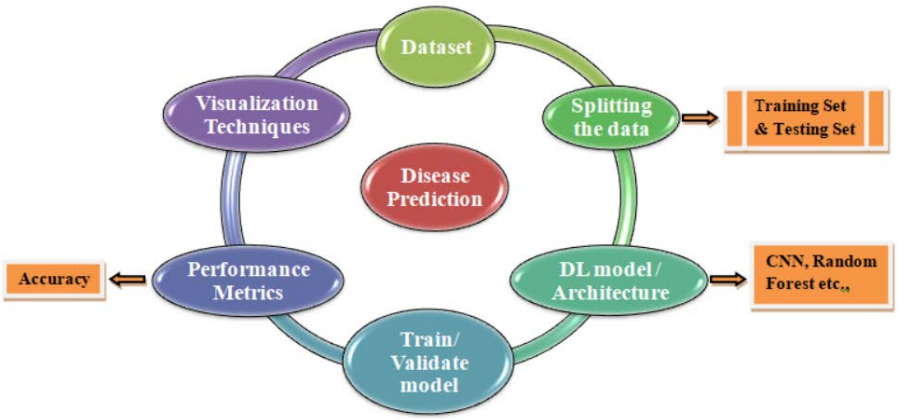
S. Roobini,* M.S. Kavitha, S. Karthik

Department of Computer Science & Engineering, SNS College of Technology, Coimbatore, Tamil Nadu

Received on: 12-Sep-2023, Accepted and Published on: 03-Jan-2024

Review

ABSTRACT



Over the recent years, conventional artificial intelligence (AI) has witnessed a significant infusion of machine learning and neural networks, marking a substantial evolution in various domains due to their autonomous capacity for feature acquisition and remarkable efficiency. Particularly in the medical field, machine learning-based models have outperformed physicians, exhibiting greater accuracy. Diseases such as cancer, Alzheimer's, dyslexia, skin diseases, and heart diseases have become focal points in medical research. Several deep learning methods, including Convolutional Neural Networks (CNN), Support Vector Machines (SVM), Random Forest, Logistic Regression, Decision Tree, and Recurrent Neural Networks (RNN), play crucial roles in disease prediction. This survey emphasizes the critical analysis of which deep learning models achieve higher accuracy in predicting specific diseases. The objective is to shed light on existing shortcomings in disease prediction and propose potential remedies for future improvements. Results indicate that Convolutional Neural Networks excel in predicting heart and Alzheimer's diseases, as well as breast cancer. Support Vector Machines demonstrate effectiveness in cancer prediction, while logistic regression proves adept at predicting dyslexia, and decision trees emerge as a favorable choice for skin diseases. Looking ahead, the integration of digital twins for predictive analytics, facilitating the simulation and modeling of disease progression based on individual patient characteristics, and leveraging blockchain for secure storage and sharing of health data represent promising avenues for future developments.

Keywords: Convolutional Neural Network (CNN), Decision Tree, Digital Twins, Disease Prediction, Logistic Regression, Machine Learning, Random Forest, Recurrent Neural Network (RNN), Support Vector Machine (SVM).

INTRODUCTION

Currently, numerous researchers employ machine learning, deep learning, and data mining techniques in the field of healthcare for the purpose of predicting diseases. However, each approach brings its own set of advantages and limitations, requiring careful consideration, but the study provides the viewpoint and prediction accuracy based on the research. The study's major goal is to develop a unique way of predicting multiple diseases using deep learning (DL). The complete prediction procedure encompasses several stages, namely: (a) acquiring the necessary data; (b) selecting the most suitable features; (c) extracting statistical features; and (d) conducting the actual prediction. The health records of various disorders are acquired from several benchmark sources in the first stage. The best statistical feature method is then useful for an accessible collection of characteristics.[1]

Of the approximately 11 million deaths that occur in India each year, 28% are attributable to cardiovascular disease. The Centres for Disease Control and Prevention estimate that heart disease claims a life in the US every 36 seconds. Cancer has one of the highest global death rates ever documented. Cancer alone would be the cause of 67 lakh fatalities in 2020. Over 10 lakh Indians lose their lives to Alzheimer's disease every year, and by 2025, the

*Corresponding Author: S. Roobini
Email: srruby13@gmail.com

Cite as: *J. Integr. Sci. Technol., 2024, 12(4), 787.*
URN:NBN:sciencein.jist.2024.v12.787

"Indian Heart Association" predicts that nearly 10 crore people will have dementia. Recent advances in machine learning in healthcare have focused on the interpretation of frameworks.[2] In most medical instances, precision is important. Despite the loss of accuracy, interpretability is frequently selected in certain situations. The industry's top need for healthcare is better illness categorization and prognosis. Disease prediction is a complicated task that an individual should be obliged to solve before it reaches a crucial stage. Individuals' risk is lessened if the condition is predicted. The expense of early-stage prediction is equally high. More than one illness can be predicted using the multi-disease prediction model. By reducing the need for extra models, this multi-disease prediction system hopes to decrease processing times and death rates. As a result, this work presents an illness prediction model created particularly for the reliable forecast of numerous diseases.

A thorough search was conducted to locate research that used deep learning and machine learning techniques to forecast multiple diseases. The PRISMA guidelines were used for this review.[3] The preferred reporting items for systematic reviews and meta-analyses, or PRISMA, have a flowchart with four phases, which include identification, screening, eligibility, and inclusion.[4] The selection criteria included only English-language publications that appeared between January 2015 and January 2023. Because of the technological and methodological differences between older research and the current standards, articles published before 2015 were not included.

A sophisticated search with the phrases "machine learning" OR "deep learning" OR "neural network" AND "disease prediction" in Scopus, Web of Science, ScienceDirect, and IEEE Xplore databases was used. Only the publications that remained after deleting duplicates were considered, which include (i) disease prediction, (ii) usage of neuroimaging data, (iii) classification techniques based on ML, DL, or NN algorithms, and (iv) accuracy findings. Following the selection of research, the information was taken based on each study: the primary author and the year of release, classes, samples taken, database, neuroimaging methodology used, features used, classification approach, and accuracy attained.

Literature Survey

Today's age is incredibly busy in our daily regular schedule, which causes anxiousness, agitation, and fretfulness. Every person has a unique pulse rate and blood pressure, with average pulse rates of 60 to 100 beats per minute (BPM) and average blood pressure readings of 120/80 to 140/90. In recent years, automatic disease detection has been a major area of research. Visibility, machine learning, or technology for their detection are typically chosen and employed. Figure 1 depicts the number of research articles reviewed for each prediction of disease.

Prediction of Heart Disease

In the medical field, especially for the prediction of heart disease, machine learning and image fusion have been detailed by Diwakar et al.[5] Utilising precise feature selection methods enhances the accuracy of the algorithm, leading to more reliable outcomes in predicting heart disease. If a specific type of heart ailment is detected, it is crucial for the patient to receive tailored treatment for that particular condition. As a result, the researchers concluded that constructing a prediction model for heart disease necessitates a dataset with dependable data and appropriate samples. The pre-processing of the dataset is essential to prepare it for use with machine learning algorithms, thereby improving the quality of the results obtained. Selecting the right algorithm is crucial to developing a prediction model that delivers accurate outcomes. The author of the study observed that artificial neural networks (ANN) had a positive impact on heart disease prediction in most models.

Sharma et al.[6] employed classification techniques such as a data set of cardiac conditions and hyperparameter optimising K-Nearest Neighbour and Support Vector Machine and examined classification accuracy. The author discovered that the best classification methods for the heart disorders dataset are hyper-parameter optimisation using Talos. Talos improves the precision of several additional improvements. To get the best forecast, it is useful to record records related to cardiac sickness.

Bharti, R. et al.[7] used three methodologies for comparison examination, and encouraging conclusions were drawn. The author revealed the performance of machine learning approaches in this investigation. According to recent research, numerous scholars have supported the utilisation of machine learning (ML) techniques, particularly when dealing with small datasets. This approach has demonstrated the advantage of reducing the overall processing time, making it favourable for deploying the model. Furthermore, it has been identified that the information within the dataset needs to be properly correlated. Failure to establish correlations can lead to overfitting of the training model, resulting in inadequate accuracy when evaluating the model with real-world data problems that significantly differ from the training dataset. Therefore, ensuring appropriate correlation is crucial to ensuring the model's effectiveness and generalizability.

Rani et al.[8] suggested an improved decision support system for identifying cardiac disease that outperforms existing methods. Whether at The authors used the finest algorithms they had found throughout recreation at the steps of missing value, selection of features, or classifier selection while creating a combination system to support decisions. The cleve land dataset was used in a Python-based simulated environment to evaluate and compare the proposed approach. The system was optimised using the randomised search CV approach.

El-Hasnony et al.[9] presented a model for effectively treating patients prior to a cardiac arrest, and it is crucial to properly forecast using machine learning techniques to study coronary artery disease algorithms. The classification machine learning strategy called "active learning" takes advantage of the extrapolation notion as opposed to the memorising concept used by traditional strategies for categorization. The results used MMC, Random, Adaptive, Quire, and AUDI are five multi-label training selection techniques that iteratively seek the most pertinent data to decrease labelling costs. In the absence of labelled data, in order to improve precision in classification and the F-measure, the grid-based searching strategy was applied. The cardiovascular disease dataset methodology uses an identification scoring algorithm as a

categorization engine. The results showed that the learning model could accurately generalise beyond the sample data.

Today, cardiac illness[10] is one of the primary causes of mortality and is a significant issue that is prevalent in developing countries, particularly in regions like Africa and Asia. Early detection of cardiac disease not only aids individuals, but also helps doctors find out the major source of the reason and also helps in predicting the severity before actual occurrences. This paper helps in proposing a method named CARDIO, which helps in analysing the likelihood of cardiac illness being present in people through machine learning algorithms and conventional neural networks. The proposed method is keen on developing a model that helps in utilising CNN for prediction at a very early stage. The experimental findings showed that the suggested technique is better in performance when compared with the existing methods using the available metrics. The results yield a high optimal accuracy of 97%.

The modern Heart attacks are one of the worst diseases and are caused by lifestyle,[11] which is common among young people. Early diagnosis helps in effective treatment among people. Hence, it is vital to develop a decision-making system with the help of available data. The information is highly useful for making predictions about health issues among the common people. Deep learning techniques are themselves highly adaptive algorithms that help in processing both structured and unstructured data effectively and yield better patterns and analyses. The research study presents a clear technique for forecasting the occurrence of heart disease being developed with the help of deep learning approaches. 14 medical features are being used in our proposed approach. A multilayer perceptron has been used in the predictive model, in which the layers are densely connected. 13 features are being used in the first layer, and the output neuron utilized for the target class is classified as binary. Many different machine learning classification techniques are implemented for the purpose of contrast, a suggested strategy that helps in estimating individuals' propensity for developing heart disease.

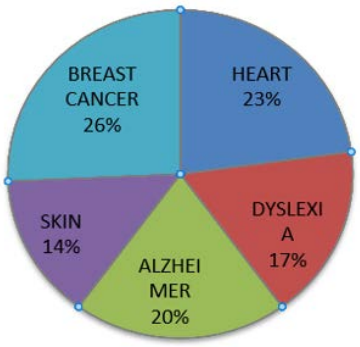


Figure 1. Number of research articles reviewed for Disease Prediction

PREDICTION OF DYSLEXIA

Dyslexia is a neurological illness that impairs the learning of individuals who have it, primarily youngsters, and causes trouble reading and writing. Dyslexia is a neurological illness that impairs the learning of individuals who have it, primarily youngsters, and causes trouble reading and writing. When dyslexia goes undetected, it causes fear and frustration in the afflicted children as well as their family members. Children who get no early assistance may enter high school having achieved considerable discrepancies. As a result, early identification and dyslexia treatment programmes for kids are critical and highly recommended in order to assist youngsters in building good self-worth and achieving their complete intellectual potential.

Raatikainen, P. et al.[12] created an eye movement classifier to detect dyslexic readers. Importantly, the author characterized the text proficiency rating used here as an artificial threshold for dyslexia and a practical approach that resulted in an intrinsically challenging categorization assignment. Their feature extraction improves on the usual by using a transformation array utilizing looking sequences inside AOIs and diagrams as opposed to averaging obsession and saccade measures. A support vector machine classifier that used the major important sight-jerking characteristics chosen by Random Forest achieved an exactness of 89.7 percent and a recall rating of 84.8 percent. An outcome that is encouraging, which is a more in-depth examination of the features of importance, gives information that may be utilized to guide future research towards rapid and reliable dyslexia screening systems.

Zahia, S., et al.[13] used a deep learning neural network to detect dyslexic children using 3D brain activity volumes. Unlike current approaches, which usually suggest a strategy that exclusively relies on brain activity zones to address the classification issue using whole-brain functional magnetic resonance imaging in two dimensions or three dimensions, The fMRI scans went through processing using SPM12 programmes to prepare the input data. This involved transforming Dicom to Niftii, modifying the movement of the head to lower inter-subject variation, normalizing to a model Atlas to lower intra-subject variability, and applying an adaptive filter to lower variability. The functional sections of the brain were then retrieved using statistical parametric maps created with the same software. The author chose to preserve the brain regions important for linguistic and text comprehension in order to decrease neurological activity in regions that might have been evident in every individual and would, thus, limit the effectiveness of feature acquisition.

Usman, O.L., et al.[14] concentrated on examining existing approaches to deep learning and machine learning for dyslexia and biomarker identification. Dyslexia, a complex brain developmental condition, has recently piqued the curiosity of researchers in contemporary neuroscience and machine learning. Despite the fact that significant Over the previous two decades, machine learning techniques have been used in this field of study; however, the use of deep learning algorithms is still in its early stages. Additionally, according to their research, SVM is the machine learning technique that is most frequently used for diagnosing and predicting dyslexia. Data for identifying and analyzing dyslexia has been acquired from a number of sources. This review believes that achieving high classification performance with CNN-based models for dyslexia and its biological indicator detection is possible if the identified potential challenges are overcome, given the growing use of deep learning in contemporary medicine and the recent modern efficacy demonstrated by the models used by convolutional networks.

Dyslexia[15] is a problem associated with the neurological system, which leads to many stumbling blocks and challenges during this procedure of learning, particularly investigating. However, the group afflicted with dyslexia may also suffer from spelling and other disabilities. Yet these do not directly pertain to their IQ. The rapid treatment will aid the kids in developing their skills with the help of specialized and well-designed software. In order to recognize this disease, the utilization of methodologies for deep learning and machine learning is widespread. through various datasets that are obtained from the organizations. This research study helps in analyzing the performance of predicting the disease and also helps in summing up the challenges faced while using these models for diagnosis and classification purposes. With the help of the PRISMA protocol, various articles have been reviewed and analyzed for the purpose of obtaining information, pre-processing, and extraction of features. This also paves the path for various researchers to build a model for this disease based upon the availability of the interrelated datasets. The study also shows us some of the challenges faced by the researcher.

Dyslexia[16] is a kind of learning disorder associated with a lack of reading and writing and, at times, poor spelling. These individuals also face eye difficulty while interpreting the words. This is a common reading disability and also a lack of expression. Children at the earliest stage experience highly discrimination and social exclusion. This is one of the widely recognized neurological disabilities that have been very common among children and affect their confidence and social-emotional development. Various researchers have shown that the earlier diagnosis helps them overcome it easily. Many research works have been approved to be understood with the help of machine learning, image processing, and also the understanding of the various behaviors of the brain with the help of psychology. In this research work, many images of the brain have been screened to understand the nature of the individuals. The research work also helps in encouraging machine learning in various external environments. The model being proposed makes use of machine learning algorithms.

PREDICTION OF CANCER

Deep learning approaches emphasize on use of an end-to-end approach.[17] Due to their excellent accuracy and interoperability with different mammography systems, deep learning algorithms have a tremendous potential to improve the precision of breast cancer detection on screening mammography. The likelihood of accuracy improving is very high when training datasets and computer resources become more widely available. This study compared and contrasted CNN, SVM, and Random Forest in order to better understand the various machine learning techniques for diagnosing breast cancer. The results showed that CNN performs better than the other techniques being used right now in terms of accuracy, precision, and data requirements. When comparing the accuracy percentages, it was found that CNN had a substantially greater accuracy of 99.67 percent compared to SVM's 89.84 percent. Accuracy rates of 90.55 percent were similar for SVM and Random Forest. These findings demonstrate CNN's remarkable accuracy and performance. According to these results, our method might prove useful for future developments in computer-aided detection (CAD) systems. These techniques might help radiologists identify the most important cases to focus on for examination after performing a first, unbiased interpretation.

Islam et al.[18] initiated a comparison analysis of five predictive machine learning techniques for diagnosing breast cancer: support vector machine [SVM], K-nearest neighbour [KNN], random forest [RF], ANN, and logistic regression. The ANN method under evaluation has a 98.57 percent accuracy rate, which is a great value. On the other hand, logistic regression (LR) and random forests (RFs) showed lesser accuracies of 95.7%. In the medical industry, the diagnosing method is both costly and utilising the aforementioned procedures for diagnosis can take some time. Machine learning approaches, however, can be helpful clinical assistants in the event of a misdiagnosis, especially for junior physicians. The model created using artificial neural networks (ANNs) is regarded as the most trustworthy of the strategies presented. It has a significant potential to revolutionise the field of breast cancer prediction.

Deep learning models may give good estimations in the detection of breast cancer.[19] According to the general performance standards, LSTM (long short-term memory) and GRU (gated recurrent unit) have proven to be highly effective in producing beneficial outcomes. One reason might be that these two algorithms have internally recognised characteristics that have a substantial influence on shaping up the training performance, resulting in better accuracy than the other algorithms utilised in this study. Several studies have already been conducted by other academics; however, their evaluations were not as thorough as ours, since using a range of performance measures, we looked at more than seven deep learning models. In addition, we produced a heat map with a substantial number of connected features and exhibited several optimisation tests to demonstrate the validity of our evaluation procedures.

Gupta, P. et al.[20] demonstrated the operation for the Wisconsin Breast Cancer Dataset. There are six machine learning models using their hyper-parameters. Deep learning and other machine learning techniques were used to perform supervised categorization of cancerous and benign cells. Adam Gradient Learning has the best accuracy since it combines the advantages of RMSProp and AdaGrad. While RMSProp thrives at non-stationary signals, AdaGrad shines at computer vision problems. In this situation, vanishing gradients were avoided by using the rectified linear unit (ReLU) function, which also made it possible for the model to train more quickly and perform better.

Sekaran et al.[21] used deep learning technology to implement FISHMAN. The early detection of abdominal malignancies like pancreatic cancer or tumours depends critically on the analysis of multiple sets of images and diagnosis based on threshold parameters and features produced using the Gaussian Mixture Model (GMM) with the Expectation-Maximisation (EM) algorithm during the training phase. These image collections are essential for spotting possible cancers at an early stage. Additionally, the comparison of the tumour stage's weight before and after the therapy time demonstrates a successful result, with a reduction in tumour size seen after the treatment. This suggests a positive

therapeutic response and possibilities for the patient's condition to get better.

Cancer[22] is a disease that makes people feel insecure and lose their confidence across the globe. But early diagnosis can help cure the disease with the proper medications. Nowadays, computer-aided diagnosis has been commonly used among people because of its effectiveness and screening, which helps save people from deadly stages. Artificial intelligence is used in the model. This research paper concentrates on developing a model by means of a deep neural network that aids in predicting breast cancer malignancy. Data from UCI is widely used, and the model yields optimised results against the other model and also shows a score of F1 over 98.

Breast cancer is a highly dreadful, risky illness with a high fatality rate.[23] An accurate diagnosis helps save the lives of many people. Practitioners started using machine learning and deep learning techniques intended for their analysis and to make important decisions. The proposed model utilises, for the diagnostic, RNN and Keras tuner optimisation approaches. An input layer, five hidden layers, five dropout layers, and an output layer made up the RNN. The number of neurons in every hidden layer can be easily optimized. To extract the feature from the data set, three feature selection techniques have been used. The various algorithms were being compared with the proposed model. The primary features were incorporated by the optimised deep RNN to determine its efficiency. It was clear from the data that the suggested model has the highest performance for CV.

PREDICTION OF ALZHEIMER DISEASE

Duc N.T. et al.[24] discuss the development of a 3-D deep learning-based computer-aided diagnosis method. It makes use of group-level ICA features from resting-state functional magnetic resonance imaging (fMRI) and machine learning regression. This method seeks to predict the clinical Mini-Mental State Examination (MMSE) score while automating the diagnosis of Alzheimer's disease. This method offers the potential for a precise and quick diagnosis of Alzheimer's disease, as well as the ability to anticipate each person's cognitive state using their MMSE score. It does this by utilising deep learning algorithms and analysing the 3-D fMRI data. The study's findings suggested that the prediction of people's MMSE scores and the prompt identification of Alzheimer's disease may both benefit from functional brain feature collection. They were able to predict participants' MMSE scores using just imaging data by using the devised and implemented machine-learning method presented. In cases where clinical professionals are few, the automated technique can serve as a scientific additional test for the early identification of cognitive decline problems.

Sun et al.[25] used a modified version of the UTMOST (Unified Test for Molecular Signatures) technique to create gene expression prediction models. The number of prediction models with at least a 0.01 performance (correlation of at least 10% between predicted and measured expression) varied among the various brain tissues from 5015 to 8582. For the spleen tissue, 8759 models were built with a performance R2 of 0.01.

Pang, Z, and colleagues[26] explored how the huge quantity of facts created by Alzheimer's disease can be utilised to aid in disease diagnosis and progression prediction. As previously stated, the focus of ADNI's multi-angle study is on verticalization, with AD serving as its central tenet. The study's goal is to concentrate on imaging data, with genetic, clinical, biochemical, and other data streams serving as supplements. Its goal is to diagnose and monitor the progression of Alzheimer's disease (AD) as soon as possible. On the basis of the ADNI datasets, they conducted a multi-modal, heterogeneous data fusion study. In this work, the multi-modal data fusion study on Alzheimer's illness is the main topic. Following are the steps being taken in the work: researching and consulting relevant literature on Alzheimer's diseases; comprehending the various data modalities that affect Alzheimer's diseases; gathering and laying out significant collections of data pertaining to help-taking decisions and disease detection; and also incorporating this initial information.

J.H. Park et al.[27] investigated the efficacy of countrywide The major goal of this study is to forecast the occurrence of Alzheimer's disease (AD) across time using population-centred operational health data. In order to predict the prevalence of AD over the course of a whole year, the researcher employed machine learning algorithms. The accuracy level of the result, which was 0.713, was satisfactory. The exceptional precision of administratively data-based prediction models, which were developed using large-scale national samples, may be relevant in the context of AD. This study demonstrates that administrative health data can be a useful source of information for AD prediction when used in conjunction with data-driven machine learning algorithms. Despite the limitations of administrative health data, such as the inability to explicitly define clinical phenotypes, it has the potential to predict AD risk.

Alatrany, A. et al.[28] sought to establish connections between regularly occurring changes. Deoxyribonucleic acid, or DNA, is genotyped as single nucleotide polymorphisms in a group of people and a phonotype. Despite the existence of numerous projects using human genomic data to predict chronic illnesses, additional study is needed. To make predictions and conduct genome-wide association studies, machine learning techniques are commonly employed. Random Forest was used in this study to identify the most significant SNPs associated with Alzheimer's illness. The illness was then predicted using a deep learning model. Their thorough simulation findings showed that using a convolutional neural network and multilayer perceptron, this hybrid model shows promise in predicting individuals with Alzheimer's disease, with areas under the curves of 0.9 and 0.93, respectively.

Early recognition of Alzheimer's disease[29] is crucial for both patients and carers in clinical trials. This research area helps in developing a novel strategy based on deep 3D conventional neural networks that aids in analysing Alzheimer disease using structural MRI. For the purpose of comparison, the reference model has been developed. Validation of both models helps in giving a clear notion of the developed model. The model developed helps in the identification of mild and severe symptoms of the disease. The proposed model focuses on the vast array of levels connected to this illness. From the results, it was clear that using neural networks, it is possible to automatically detect the biomarkers that are associated with predicting diseases.

An overview study of Alzheimer disease started out using two machine learning methods that aid in identifying disease at the earliest stage.[30] This disease can occur at any age and tends to affect 46.8 million people across the world. This directly has an influence on a person's life. This disease is expected to increase to 1.68 million by 2050. Many studies have been carried out on this disease, which slows down the neurons in the brain. This disease can also lead to unusual behaviour and language problems in people. The tangle bundles of neurofibrillary fibres in the brain are the main source of this disease. The initial stage has to be keenly analysed for easy recovery and to prevent further losses. The higher symptoms might include a loss of memory, wrong judgement, difficulty identifying the objects, confusion in day-to-day activities like driving, and also placing the objects in different places. And this has been divided into three periods: primary, intermediate, and the last period of dementia. There are many treatments to save people.

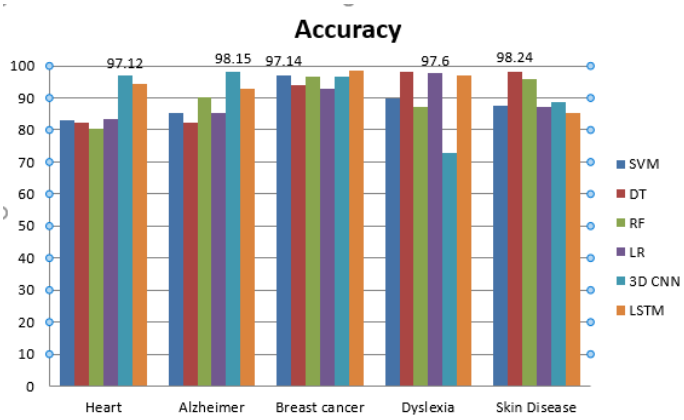


Figure 2. Accuracy comparison of various models for Disease Prediction

PREDICTION OF SKIN DISEASE

Srinivasu et al.[31] The MobileNet V2 and LSTM techniques were used to create a model that successfully identified and classified skin disorders while utilising little computational power and time. The simulation's analysis yielded an intriguing score of 85.34 percent while compared to other algorithms using live photos from Kaggle. To work with a transportable Stride2 device, the MobileNet V2 framework was selected. The forecast accuracy might be increased by using the MobileNet V2 architecture, maximising the model's computing performance, and keeping past timing data using LSTM. By maximising weight while taking the current situation into consideration, the model's endurance may be improved. Furthermore, this system was contrasted with other well-known models, including CNN, FTNN, and HARIS. The recommended approach excelled others in classifying tumours and determining their trajectory through development by using data that was textured-based.

Li et al.[32] discussed the evolution of findings from computational research in the field of dermatological illness detection as well as diagnostic technology, the procedure of conventional healthcare assessment, and machine learning-based image identification of skin disease.

Accurate detection is a time-consuming operation, which raises the necessity for a dependable automated detection technique that professional and non-expert doctors may use consistently in the diagnostic process.[33] Deep learning is a broad field that necessitates extensive understanding in engineering, information, computer science, and medicine. With the continuing expansion of the aforementioned sectors, deep learning is undergoing rapid development and has piqued the interest of many countries. Deep learning for skin disease diagnosis is clearly a promising approach in the near future, fueled by more inexpensive solutions, software that can swiftly collect and usefully interpret large data, and hardware that can perform what people cannot.

Dermatology[34] has become one of the most challenging fields, as there are many complexities involved in it. It is very common to do regular practice to help individuals determine the type of skin condition they suffer from at various times, and many techniques are practiced by various doctors to analyse the problem associated with the skin disease. This research work provides an automated method based on the image that helps in diagnosing and categorising the problems associated with the skin through machine learning classification. A computational approach helps to analyse processes and segregate the picture. Skin photographs are initially filtered in order to eliminate unwanted processing, remove image noise, and improve the quality. Extraction of the feature from the image is achieved through CNN, and the picture can be easily classified through an inspection report using the Softmax classifier algorithm. This application is highly helpful and efficacious for people to diagnose skin diseases.

Skin disease[35] has become common among millions of people. Usually, skin diseases have many hidden dangers that lead to a lack of self-confidence and psychological depression. At times, it even leads to a higher risk of skin cancer. Many experts are trying to diagnose the diseases and find a remedy at an earlier stage. The proposed framework makes use of techniques such as CNN architecture and other models called AlexNet, Resnet, and Inception V3. The data set was expanded with the help of other images, which can then be classified easily. The deep learning algorithm helps with quick analysis by reducing human labour.

Doctors can treat and care for patients more effectively by using predictive models based on machine learning algorithms, which can assist in the accurate and timely detection of diseases. It is a fantastic idea to use machine learning algorithms to detect various ailments, including cancer, dyslexia, Alzheimer's, heart disease, and skin conditions. The prediction model's overall efficacy can be increased, and maximal accuracy can be attained by utilising algorithms like random forest, decision tree, logistic regression, 3D CNN, and so on. To guarantee patient safety and wellbeing, it is crucial to verify the model's accuracy with real-world data and to have the results validated by a medical professional. From the survey, it has been reviewed as 3D-Convolutional Neural Network (3D-CNN) works well for heart and Alzheimer as their accuracies are 97.12% and 98.15%, then cancer is trained well with Support Vector Machine (SVM) with its precision of 97.14%, 3D-CNN is noted as 96.5%, dyslexia performs well for Logistic Regression

(LR) with 97.6%, and finally skin disease well fits with Decision Tree (DT) as 98.24% is depicted in figure 2 based on accuracy metrics.

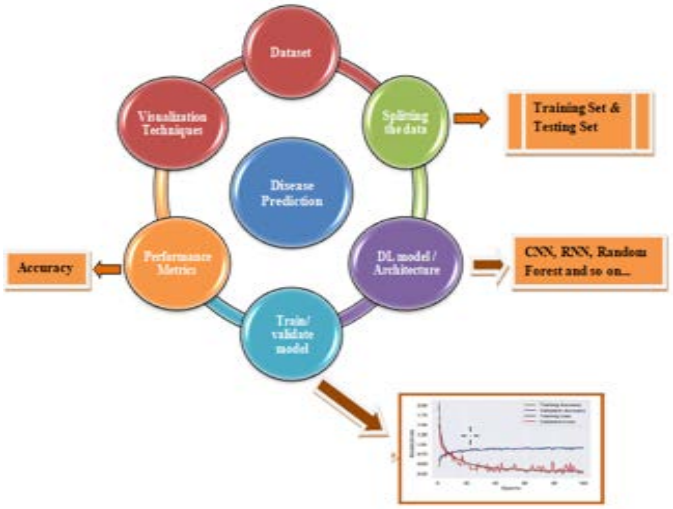


Figure 3. Prediction Cycle Model

Deep Learning Algorithms

Deep learning algorithms play a crucial role in our lives, resulting in outstanding results on several cognitive tasks matching or even growing those provided by human performance. The greatest advantage of DL is its capacity to learn huge amounts of data. The DL field has grown rapidly in the last few years and has achieved success in a wide range of applications. DL has outperformed well in many domains like cyber security, natural language processing, bioinformatics, robotics, and medical information processing. Figure 3 depicts the standard prediction model for any disease. The impact is already seen in cancer, diagnosis, precision medicine, forecasting, speech recognition, etc. In large databases, it's highly difficult to extract needed features with existing methods. Deep learning algorithms use highly optimised algorithms and architectures. These are a few of the widely used deep learning algorithms:

Convolutional Neural Network

CNN is a component of ANN in deep learning, which is frequently used for visual imagination. It is also known as space-invariant ANN, which is based on the architecture. CNNs are categorised versions of multilayer perceptrons. Multilayer perceptrons generally imply networks that are interconnected, where each neuron is linked to the neurons in its immediate layer. The idea of complete connectedness in networks makes them susceptible to data overfitting. CNN takes advantage of hierarchical patterns in the data as well as assembles patterns of complex increase with the help of smaller and simpler patterns in their own filters.[13] As a result, the connectivity is at its lowest point. The connection pattern between neurons that underlies the structure of the visual cortex in biological processes is another source of inspiration for these networks. Comparatively speaking to other image classification techniques, CNN's needs less pre-processing. This suggests that these networks use automatic learning to attempt to optimise the kernels, as opposed to traditional methods where they are hand-engineered.

Recurrent Neural Network

RNN belongs to an art algorithm in which sequential data is all used for voice searches on Google and Siri on Apple. This method is the first to recall its input since it has internal memory, which makes it appropriate for machine learning issues involving sequential data.[18] The RNN algorithm has stood as a backbone of deep learning algorithms over the past few years. This is a powerful and robust type of neural network. RNN finds it challenging to learn long-distance information since the gradient in the network will either get extremely small and eventually vanish, or it will get too big. The long short-term memory (LSTM) network was presented as a solution to this issue. It can store sequence data for an extended period of time and resolve the gradient disappearing issue.

Support Vector Machine (SVM)

A finite-dimensional vector space with dimensions for each object feature is referred to as a support vector, which is a supervised learning technique. The consistency of the dataset and the pre-processing methods had an impact on the predictive model's accuracy and efficiency, in addition to the algorithms that were employed. Errors, missing data, redundancies, noise, and a host of other problems can make a dataset unsuitable for use with machine learning algorithms. The size of the dataset is another factor.[7] Studies involving both regression and classification make use of this algorithm. Data is plotted in n-dimensional space using coordinates, and there are two types of SVM: linear and nonlinear. The linear SVM classifier will be used because the data is linearly separable. The best hyperplane is identified in order to classify the classes, and the hyperplanes are the lines that separate the classes into groups, as depicted in figure 4. In two dimensions, the line is a hyperplane to divide the classes.

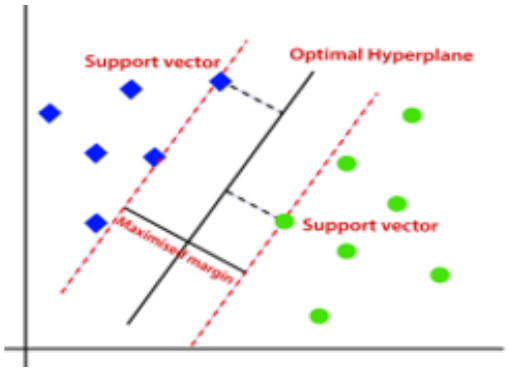


Figure 4. Support Vector Machine (Svm)

Decision Tree

The decision tree family of algorithms, which is frequently employed in supervised learning, is the Decision Tree Classifier, a machine learning algorithm created to address classification problems. The way the algorithm works is that it builds a structure resembling a tree, with each node representing a decision based on the values of multiple attributes. The path from the root node to the leaf nodes is used by the tree to make its final decision. Choosing

the feature with the highest information gain is a measure of how much a feature contributes to lowering uncertainty about the target variable. This is the first step in the decision tree classifier process.[40] The data is divided according to the values of this chosen feature, which serves as the root of the tree. Iteratively partitioning a node continues until a stopping condition is met, such as hitting the maximum tree depth or a predefined threshold for the number of samples in each node. The decisions made by the algorithm and the connection between the features and the target variable are depicted as a tree diagram. Large datasets can be handled by the algorithm because it can handle both continuous and categorical data.

LOGISTIC REGRESSION

One of the machine learning classification algorithms called logistic regression is used to evaluate datasets that contain both categorical dependent variables and one or more independent variables that determine an outcome. While logistic regression converts its output using the logistic sigmoid function to return a probability value that can then be mapped to two or more discrete classes, linear regression uses its output in continuous numerical form. Rather than using a linear function, the logistic regression model makes use of a more sophisticated cost function called the sigmoid function, or logistic function.[41-43] The cost function is constrained by logistic regression between 0 and 1.

RANDOM FOREST

The foundation of Random Forest is collective learning, or the synthesis of numerous classifiers. A collection of distinct silhouettes is called an ensemble. Ensemble employs two techniques:

- The first random set of data is used to start the bagging process. After that, it is divided into bootstrap samples; this procedure is called bootstrapping. Additionally, each model is trained separately to produce distinct outcomes, known as aggregation.[17] All of the results are combined in the end process, and a majority vote determines the outcome.
- Boosting: Boosting turns ineffective learners into proficient ones by building a series of models that culminate in the highest-accuracy final model. Because the conclusion in this case is based on the average or overwhelming rating and the predictive models are derived from subsets of data, the risk of overfitting is decreased.

Entropy is the unit of measurement for an object's disorder. Selecting a root node with low entropy is the way the Random Forest Algorithm builds trees.

COMPARATIVE STUDY OF MACHINE LEARNING MODELS FOR VARIOUS DISEASES

Artificial intelligence and neural networks play a major role in health care, especially in predictive analytics. In health care, accuracy is an important metric to be analysed for the developed model. The model developed is evaluated for different kinds of datasets for a particular disease or different datasets for different diseases. Foremost, the researchers need to understand the impact of analysing the model for disease prediction. For a specific dataset, different algorithmic models can be used to evaluate it, which is considered worthwhile.

In this section, for a particular disease, many deep learning, machine learning, and neural network models are developed and evaluated by considering accuracy as a performance metric. Based on the performance, the specific model works fine for that kind of particular disease. In the research area, the algorithms are compared to determine their higher accuracy in predicting dyslexia, cancer, Alzheimer's, and so on. Table I provides a summary of all algorithms for each disease prediction and make a comparison of each technique with its accuracy, findings, and limitations.

Table 1. Summarizes the relative study of numerous algorithms for disease prediction

Algorithm used	Findings	Limitations	Acc (%)
DYSLEXIA			
Support Vector Machine [SVM] [12]	SVM predictor that analyses eye movement data to determine dyslexic readers	No fast and reliable screening tools	89.7
3D Convolutional Neural Network combined with dense layer [13]	A deep learning neural network to identify dyslexic children from 3D brain activation volumes	Challenges in virtual inputs of kids	72.73
Decision Tree [16]	Difficult in structuring the images.	Adaptation of the external environments	97.61
Particle Swarm Optimization (PSO) based Hybrid Kernel SVM-PSO [37]	Improved accuracy, sensitivity and specificity over Linear SVM	Feature extraction technique is not efficient.	97.23
SKIN DISEASE			
MobileNetV2 with LSTM [33]	Deep learning classification of skin diseases	Application in real time varying data	85.34
Softmax classifier and Convolutional Neural Network [34]	More efficient than the conventional method. - Deep learning classification of skin diseases.	analysis of the complex image texture	87.20
RESNET152V2 [35]	Classification of the skin disease using predefined models.	CNN outperformed on training data and not testing data	88.83
HEART DISEASE			
ANN [5]	Heart illness using machine learning and image fusion.	Lack of specific feature selection. Not suitable for lively database	88.43

DNN with Chi-Square statistical model [6]	Hyper-parameter optimization on Heart diseases for better accuracy	Optimal results could not achieve as dataset was larger	90.2
Comparison - Logistic Regression, KNN, SVM, Random Forest, Decision Tree and Deep Learning [7]	Heart diagnosis using machine learning and deep learning with reduced computational time in which deep learning Technique provides high accuracy.	Using real world dataset is not efficient	94.20
Genetic algorithm with Random Forest [8]	Randomized Search CV for accurate prediction of heart disease	Severity and chronic diseases is not accessed	86.67
Convolutional Neural Network [10]	Optimal results with high metric values.	Not applicable to the chronic disease.	97.12
Multi Layer Perceptron [11]	Yielding high end patterns and analysis	MLP suffers from over fitting, under fitting. -high time consumption	92.45
BREAST CANCER			
CNN ,SVM and random forest [17]	Machine learning approaches for spotting breast cancer. CNN provides better accuracy.	Imaging issues might arise in medical complications	95.90
Deep neural network with Support Value (DNNSV) , Recurrent Neural Network (RNN) [18]	Five machine learning techniques—support vector machine, K-nearest neighbors, random forests, DNNS, and logistic regression have been compared for the prediction of breast cancer.	Because of the small size of the dataset, it needs to be enlarged using data augmentation.	97.14
LSTM, Radial Basis Function and GRU [19]	Deep Learning Models for Detecting Breast Cancer	Not intact with changing real, time data	98.67
K-NN, Support vector machines, decision trees, random forests, logistic regression, and Adam Gradient Descent Learning [20]	fAdaGrad and RMSProp properties are achieved	Intense medical images are not considered	98.24
Decision Tree J48 algorithm [38]	Prediction of breast cancer using machine learning algorithm	Due to the fact that data mining was utilized to get the information, more recent prognostic elements must be investigated and incorporated to the decision tree.	94.56
k-Nearest Neighbor classifier (k-NN) [39]	To detect, combining biosensors and machine learning algorithms to detect breast cancer have the ability to identify cancer faster.	The investigation's failure to employ a thermo-graphic database to retain and compare typical and atypical photos has resulted in less-than-ideal findings.	94.44
ALZHEIMER			
3-D CNN, LASSO; Recursive feature removal and a brief mental state examination [24]	3-D computer-aided diagnosis for Alzheimer's disease prediction using deep learning.	Lack of human prediction with results	90.04
Random Forest [26]	Random Forest for Alzheimer's disease	More investigations are required	90.02
Logistic regression [27]	Prediction of AD is scalable and compatible	Risk associated with early detection	85.27
Model based on 3D deep convolution neural networks [28]	Accurate premature illness recognition.	Other information like age, genetical data etc are not considered.	93.23
CNN with multilayer of 18 and 3D CNN [30]	Early prediction with help of deep learning.	More complex in analyzing the image of old people	98.15

CONCLUSION

This paper's goal is to identify methods for diagnosing diseases based on machine learning and deep learning, with a focus on diseases like Alzheimer's disease, dyslexia, heart disease, and breast cancer that are discussed in relation to these methods. Furthermore, machine learning-based disease diagnosis techniques include decision trees, random forests, recurrent neural networks, convolutional neural networks, and logistic regression. Other ML-based illness diagnostic techniques are also addressed using a deep learning technique termed a neural network. Kaggle, OASIS, ADNI, and UCI repositories were taken into consideration for forecasting the disease. The overall system results show that the convolutional neural network model performs better than the other models and offers higher disease detection accuracy. The training and testing parameters can also be increased to further improve the model. The survey's executive summary showed numerous machine learning and neural network techniques for disease prediction. This review can educate pertinent researchers, assist them in comprehending the existing models and projected growth methods for predicting diseases, and help them design more effective, efficient, and prudent research aimed at increasing medical development. Additionally, performance may be enhanced in the future with larger datasets.

FUTURE PERSPECTIVES

Digital twins and deep learning algorithms can be used together to forecast diseases and provide automated medical treatments that are quicker and more advanced. The workload of doctors will be significantly reduced in the future due to the more intelligent processing modes brought about by digital twins and deep learning models. Digital twins can be incorporated with medical health to improve dynamic monitoring, allowing for continuous monitoring of patients health status, enabling simulation and modeling of disease progression, and then personalized treatment plans. Therefore, combining deep learning and machine learning with blockchain intelligence provides a potential research area intended for developing secure diagnostic systems. Blockchain allows patients to manage and control access to their health data through smart contracts. Patients can grant or revoke access to their data, ensuring privacy and compliance with regulations. To achieve extremely adapted medical care in intelligent healthcare facilities, incorporate accuracy, and provide patients with added convenience, digital twins should be used in relation to future medical technology. Medical procedures, including illness recognition, early disease diagnosis, and treatment for diseases, may all be accomplished with the help of digital twins, offering ground-breaking healthcare solutions.

CONFLICT OF INTEREST STATEMENT

Authors do not have any conflict of interest to declare.

REFERENCES AND NOTES

1. K. Arumugam, M. Naved, P.P. Shinde, et al. Multiple disease prediction using Machine learning algorithms. In *Materials Today: Proceedings*; **2023**; Vol. 80, pp 3682–3685.
2. D. Moher, A. Liberati, J. Tetzlaff, D.G. Altman. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J. Clin. Epidemiol.* **2009**, 62 (10), 1006–1012.
3. J. Zhang, L. Han, L. Shields, J. Tian, J. Wang. A PRISMA assessment of the reporting quality of systematic reviews of nursing published in the Cochrane Library and paper-based journals. *Med. (United States)* **2019**, 98 (49).
4. M. Diwakar, A. Tripathi, K. Joshi, et al. Latest trends on heart disease prediction using machine learning and image fusion. *Mater. Today Proc.* **2020**, 37 (Part 2), 3213–3218.
5. P. Rani, R. Kumar, N.M.O.S. Ahmed, A. Jain. A decision support system for heart disease prediction based upon machine learning. *J. Reliab. Intell. Environ.* **2021**, 7 (3), 263–275.
6. I.M. El-Hasnony, O.M. Elzeki, A. Alshehri, H. Salem. Multi-Label Active Learning-Based Machine Learning Model for Heart Disease Prediction. *Sensors* **2022**, 22 (3), 1184.
7. A. Sharma, T. Pal, V. Jaiswal. Heart disease prediction using convolutional neural network. *Cardiovasc. Coron. Artery Imaging Vol. 1* **2021**, 245–272.
8. H. Tiwari. Early prediction of heart disease using deep learning approach; Deep learning for Medical Approach with Unique Data, **2022**.
9. P. Raatikainen, J. Hautala, O. Loberg, et al. Detection of developmental dyslexia with machine learning using eye movement data. *Array* **2021**, 12, 100087.
10. O.L. Usman, R.C. Muniyandi, K. Omar, M. Mohamad. Advance Machine Learning Methods for Dyslexia Biomarker Detection: A Review of Implementation Details and Challenges. *IEEE Access*. **2021**, 36879–36897.
11. N.D. Alqahtani, B. Alzahrani, M.S. Ramzan. Deep Learning Applications for Dyslexia Prediction. *Appl. Sci.* **2023**, 13 (5), 2804.
12. V.R. Allugunti. Breast cancer detection based on thermographic images using machine learning and deep learning algorithms. *Int. J. Eng. Comput. Sci.* **2022**, 4 (1), 49–56.
13. M.M. Islam, M.R. Haque, H. Iqbal, et al. Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques. *SN Comput. Sci.* **2020**, 1 (5), 290.
14. P. Gupta, S. Garg. Breast Cancer Prediction using varying Parameters of Machine Learning Models. *Procedia Comput. Sci.* **2020**, 171, 593–601.
15. K. Sekaran, P. Chandana, N.M. Krishna, S. Kadry. Deep learning convolutional neural network (CNN) With Gaussian mixture model for predicting pancreatic cancer. *Multimed. Tools Appl.* **2020**, 79 (15–16), 10233–10247.
16. S.S. Prakash, K. Visakha. Breast Cancer Malignancy Prediction Using Deep Learning Neural Networks. In *Proceedings of the 2nd International Conference on Inventive Research in Computing Applications, ICIRCA 2020*; **2020**; pp 88–92.
17. N.T. Duc, S. Ryu, M.N.I. Qureshi, et al. 3D-Deep Learning Based Automatic Diagnosis of Alzheimer's Disease with Joint MMSE Prediction Using Resting-State fMRI. *Neuroinformatics* **2020**, 18 (1), 71–86.
18. Y. Sun, J. Zhu, D. Zhou, et al. A transcriptome-wide association study of Alzheimer's disease using prediction models of relevant tissues identifies novel candidate susceptibility genes. *Genome Med.* **2021**, 13 (1), 13 1 ,1–11.
19. Z. Pang, X. Wang, X. Wang, et al. A Multi-modal Data Platform for Diagnosis and Prediction of Alzheimer's Disease Using Machine Learning Methods. *Mob. Networks Appl.* **2021**, 26 (6), 2341–2352.
20. J.H. Park, H.E. Cho, J.H. Kim, et al. Machine learning prediction of incidence of Alzheimer's disease using large-scale administrative health data. *npj Digit. Med.* **2020**, 3 (1), 1–7.
21. A. Alatrany, A. Hussain, J. Mustafina, D. Al-Jumeily. A Novel Hybrid Machine Learning Approach Using Deep Learning for the Prediction of Alzheimer Disease Using Genome Data. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer, Cham, **2021**; Vol. 12838 LNAI, pp 253–266.
22. V. Patil, M. Madgi, A. Kiran. Early prediction of Alzheimer's disease using convolutional neural network: a review. *Egypt. J. Neurol. Psychiatry Neurosurg.* **2022**, 58 (1), 130.
23. P.N. Srinivasu, J.G. Sivasai, M.F. Ijaz, et al. Classification of skin disease using deep learning neural networks with mobilenet v2 and lstm. *Sensors* **2021**, 21 (8), 2852.
24. L.F. Li, X. Wang, W.J. Hu, et al. Deep Learning in Skin Disease Image Recognition: A Review. *IEEE Access* **2020**, 8, 208264–208280.
25. B. Zhang, X. Zhou, Y. Luo, et al. Opportunities and Challenges: Classification of Skin Disease Based on Deep Learning. *Chinese J. Mech. Eng. (English Ed.* **2021**, 34 (1), 34 1 ,1–14.
26. M. Abishek, B.N.H. Kumar. Project on Dyslexia prediction using machine learning algorithms. *Int. Res. J. Mod. Eng. Technol. Sci.* 4 (9), 560–563.
27. A. Higa. Diagnosis of Breast Cancer using Decision Tree and Artificial Neural Network Algorithms. *Int. J. Comput. Appl. Technol. Res.* **2018**, 7 (1), 23–27.
28. Y. Amethiya, P. Pipariya, S. Patel, M. Shah. Comparative analysis of breast cancer detection using machine learning and biosensors. *Intell. Med.* **2022**, 2 (2), 69–81.
29. E. Agbemade. Predicting Heart Disease using Tree-based Model. *Data Sci. Data Min.* **2023**. University of Central Florida.
30. A. G, B. Ganesh, A. Ganesh, et al. Logistic regression technique for prediction of cardiovascular disease. *Glob. Transitions Proc.* **2022**, 3 (1), 127–130.
31. I. Mohit, K. Santhosh Kumar, U.A. Kumar Reddy, B. Suresh Kumar. An Approach to detect multiple diseases using machine learning algorithm. *J. Phys. Conf. Ser.* **2021**, 2089 (1), 012009.
32. S. Sharma, M. Parmar. Heart Diseases Prediction using Deep Learning Neural Network Model. *Int. J. Innov. Technol. Explor. Eng.* **2020**, 9 (3), 2244–2248.
33. R. Bharti, A. Khamparia, M. Shabaz, et al. Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning. *Computational Intelligence and Neuroscience*. **2021**, 1-11.
34. S. Zahia, B. Garcia-Zapirain, I. Saralegui, B. Fernandez-Ruanova. Dyslexia detection using 3D convolutional neural networks and functional magnetic

resonance imaging. *Computer Methods and Programs in Biomedicine*. **2020**.
35. A. Jothi Prabha, R. Bhargavi. Prediction of Dyslexia Using Machine Learning—A Research Travelogue. *Lect. Notes Electr. Eng.* **2019**, 556, 23–34.
36. P. Ghosh, S. Azam, K.M. Hasib, et al. A performance based study on deep learning algorithms in the effective prediction of breast cancer. In ***2021*** *International Joint Conference on Neural Networks (IJCNN), IEEE Xplore*; pp 1–8.
37. H. Saleh, S.F. Abd-el ghany, H. Alyami, W. Alosaimi. Predicting Breast Cancer Based on Optimized Deep Learning Approach. *Comput. Intell. Neurosci.* **2022**, 2022, 1–11.
38. S. Liu, A. V. Masurkar, H. Rusinek, et al. Generalizable deep learning model for early Alzheimer's disease detection from structural MRIs. *Sci. Rep.* **2022**, 12 (1), 12,17106.
39. T. Parashar, K. Joshi, R.R. N, et al. Skin Disease Detection using Deep Learning. In *2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART)*; IEEE, **2022**; pp 1380–1384.
40. S. Inthiyaz, B.R. Altahan, S.H. Ahammad, et al. Skin disease detection using deep learning. *Adv. Eng. Softw.* **2023**, 175 (ue VIII), 103361.
41. A. G, B. Ganesh, A. Ganesh, et al. Logistic regression technique for prediction of cardiovascular disease. *Glob. Transitions Proc.* **2022**, 3 (1), 127–130.
42. Y. Parikh, H. Koringa. Left Ventricle Segmentation using Bidirectional Convolution Dense Unet. *J. Integr. Sci. Technol.* **2023**, 11 (1), 417.
43. B. Mehta, V. Dave. Detection of Cardiovascular Autonomic Neuropathy using Machine Learning Algorithms. *J. Integr. Sci. Technol.* **2023**, 11 (4), 562.