OPEN ACCESS

Kashf Journal of Multidisciplinary Research
Vol: 01 - Issue 12 (2024)
P-ISSN: 3007-1992 E-ISSN: 3007-200X
https://kjmr.com.pk

Multi-Disease Prediction Using Machine Learning and Deep Learning Models

Anwar Ul Haq
Department of Computer Science and Information Technology, Faculty of Computing, University of Malakand, Khyber Pakhtunkhwa Pakistan
Muhammad Yaseen
Department of Computer Science and Information Technology, Faculty of Computing, University of Malakand, Khyber Pakhtunkhwa Pakistan
Abdullah khan
Department of Computer Science and Information Technology, Faculty of Computing, University of Malakand, Khyber Pakhtunkhwa Pakistan
Fakhrud Din
Department of Computer Science and Information Technology, Faculty of Computing, University of Malakand, Khyber Pakhtunkhwa Pakistan
Asad Ali

Faculty of Computer Science and Information Technology, Sarhad University of Science & Technology, Khyber Pakhtunkhwa, Pakistan

Corresponding author: Fakhrud Din fakhruddin@uom.edu.pk
DOI: *https://doi.org/10.71146/kjmr176*

Article Info

Abstract

In today's world, a large section of the human population suffers from treatable diseases such as heart problems, diabetes, skin cancer, stroke, liver disease, Parkinson's, malaria, and brain tumors. But due to lack of accessible and affordable healthcare the conditions may not be accurately or timely diagnosed that may lead to severe consequences, including disability or even death. Accuracy, affordability, timeliness and accessibility are critical factors that impact diagnosis quality. To improve diagnosis accuracy, affordability, accessibility and eliminate biases, machine learning and deep learning-based algorithms are gaining attraction in revolutionizing the healthcare industry. The objective of this study is to create a flexible and comprehensive medical diagnostic framework based on machine learning and deep learning models that can predict many diseases based on a patient's health records. The main objective of this study is to prevent problems caused by misdiagnosis and delayed diagnosis. By analyzing multiple diseases using a single platform, the cost of patient treatment can be reduced significantly, making it more accessible for people in underprivileged regions. By increasing the accuracy and speed of disease prediction, the proposed machine learning and deep learning-based diagnosis system has the potential to save lives. Experimental results showed that Random Forest has outperformed competing models on numerical datasets while on image datasets, VGG16 generated best accuracy than ResNet50..

Keywords: *Classification, augmentation, feature selection, Random Forests, SVM Classifier, ANN, ResNet50, VGG16.*

1. Introduction

In today's world a plethora of diseases such as heart diseases, diabetes, skin cancer, stroke, liver disease, Parkinson's, malaria and brain tumor are prevailing, which, in fact, are devastating to human beings. Gaining a comprehensive understanding of the historical landscape of these diseases offers valuable insights into their prevalence, impact, and the imperative of precise diagnosis. The following are some pertinent statistical data and noteworthy facts regarding several prevalent diseases.

In 2017, heart disease emerged as the leading global cause of death, accounting for 16.4 million fatalities (28.8% of all deaths), while diabetes ranked fourth for years lived with disability (YLDs) at 3.0%, stroke ranked third for YLDs at 4.6% and second for deaths at 11.3%, liver disease stood as the eleventh largest cause of death with 1.3 million fatalities (2.3% of all deaths), and brain tumors ranked seventh for YLDs with 23.5 million worldwide[1]. In 2012, nonmelanoma skin cancer affected approximately 3.3 million individuals, with an incidence rate of 1,760 cases per 100,000 people, showing higher prevalence among men and increasing with age, peaking in the 80s and beyond; basal cell carcinoma accounted for about 80% of cases, making it the most common type, followed by squamous cell carcinoma at 16%[2]. Parkinson's disease, a rapidly advancing neurological ailment, experienced a 22.8% increase in occurrences from 1990 to 2016, impacting millions worldwide, predominantly men, and showing higher prevalence rates in North America, Europe, and Australasia; It accounted for 1.6 million years of disability-adjusted life in 2016, making it the 37th largest worldwide burden [3].

Between 2000 and 2010, global malaria deaths decreased by 25%, with Africa experiencing the most significant decline, and in 2010, there were an estimated 1.24 million malaria-related deaths worldwide, with 85% of these fatalities occurring in sub-Saharan Africa; the study highlighted that young children and pregnant women were particularly vulnerable to malaria [4]. Although diagnosing these diseases at early stages is very important for saving lives, the high cost of diagnosis and unavailability of trained professional may leave some patients misdiagnosed or undiagnosed.

The failure to properly understand diseases mentioned above is a crucial matter within the health care system and the possible outcomes may include extreme suffering, disability or even death. There are different causes for failure to a diagnose the sickness correctly. One of these is the precision of diagnosis which may depend on lack of necessary equipment or trained professionals because some of the signs can be confused with others suggesting wrong course of treatment that in the end worsens the situation. Price can also be a hindrance in getting accurate diagnosis for example some patients are unable to pay for essential medical examination or advice, thus leaving the diagnosis partial and the treatment unsatisfactory. Time is important too as the increasing gap between the onset of signs and their examination forces more complications and lack of equipment or prerequisite number of skilled medical staff can delay diagnosis. More so, the belief or bias on a patient's age, sex, skin color or social class may lead to poor assessment or lack of it completely. These are the issues that need to be addressed to improve accuracy, affordability, accessibility and eliminate biasness in medical diagnosis.

This paper aims to design and develop robust machine learning and deep learning frameworks to address to solve inefficiencies associated with current systems for disease prediction. Eight 8 different diseases for diagnosis are selected carefully of which 5 are based on numerical datasets and 3 are based on image datasets. To diagnose diseases using numerical data, three models are designed. Of these, two are machine learning models and one is deep learning model. Disease diagnosis by using image datasets, VGG16 and ResNet50 models are hybridized with transfer learning techniques. The models have been trained and tested on the datasets adopted. Comparative analysis of the models' predictions accuracy is presented.

The lay out of the rest of the paper is as follows. Section 2 presents related work, section 3

elaborates adopted machine learning and deep learning models, section 4 presents the proposed framework, section 5 presents results and discussion and finally section 6 concludes the work.

2. Related work

Several articles have applied various techniques and algorithms for disease prediction that a healthcare facility may utilize. This part presents a brief overview of procedures applied adopted in the literature for disease prediction and diagnosis. Several researchers have been investigating the use of a variety of diseases prediction through a variety of machine learning approaches. For the prediction of diabetes, Ambesange [5] used sensors data and results revealed that KNN achieved highest 81.91% accuracy. Within the scope of the studies that have been conducted, Yaganteeswarudu [6] noted the comparative effectiveness of Decision Trees, Random Forest, and logistic regression algorithms and found that SVM was more accurate for diagnosis and prediction of various forms of cancer at 96% accuracy rate while Random Forest was more accurate for prediction of heart disease at 95% accuracy. Evaluating symptoms, selecting algorithms and diagnosing of a disease was carried out by Gram Purohit [7] and comparative analysis showed 95.12% accuracy rate by Decision Tree Classifier (DTC). Using a variety of visualization techniques, Shaikh [5] chose the J48 algorithm, which had the best accuracy rate of 98.12% for classification. The Random Forest Algorithm was employed by Saboji et al. [6] in an effort to discover an adaptable system that may predict heart disease through classification mining. This system evaluates it against Naive-Bayes classifiers; Nevertheless, Random Forest generates results that are 98% accurate. Making use of machine learning applications and methodologies, Kohli et al. [7] proposed disease prediction by means of methods like Logistic Regression, Decision Tree, Support Vector Machine, Random Forest, and Adaptive Boosting. This paper focuses on the prognosis of diabetes, breast cancer, and heart disease. Logistic regression yields the highest accuracy rates, which are 95.71% for breast cancer, 84.42% for diabetes, and 87.12% for heart disease. Utilizing distributed machine learning classifiers, Jena et al. [8] focused on risk prediction for chronic diseases by making the use of methods like Naive Bayes and Multilayer Perceptron. The accuracy of Naive Bayes and Multilayer Perceptron, which are used in this paper to predict Chronic Kidney Disease, is 95% and 99.7%, respectively. A system that provides better results for disease prediction was created by Chetty et al. [9], using a fuzzy approach. and employed methods such as fuzzy c-means clustering, fuzzy KNN classifier, and fuzzy KNN classifier. In this study, the accuracy of the predictions for diabetic disease and liver disorder is 97.02% and 96.13, respectively.

A number of studies in the area of deep learning have concentrated on employing convolutional neural networks (CNNs) to increase the precision of disease detection. According to Ambekar et al. [10], who suggested illness risk prediction and employed the CNN-UDRP, Naive Bayes, and KNN algorithms, Naive Bayes had an accuracy rate of 82%. In order to increase precision and effectiveness, Gram Purohit et al. [11] examined the results of the CNN model and VGG-16 architecture on MRI images of brain tumors. Revathy et al. [12] suggested a machine learning model based on CNN to detect malaria in blood smears, while Ali et al. [13] proposed a DCNN model for precise categorization of benign and malignant skin lesions. To increase accuracy, preprocessing methods like noise removal and feature extraction are performed. In comparison to other transfer learning models, the DCNN model performs best, with training and testing accuracy of 93.16% and 91.93%, respectively.

3. Adopted Machine Learning and Deep Learning Models

In this study we used machine learning and deep learning models. For numerical data we used Random Forest, Extreme Gradient Boosting and Artificial Neural Network. For images data we

used ResNet50 and VGG16. Django web app is created for deployment.

3.1 Random Forest

The Random Forest algorithm is one of the best ensemble classification methods [14]. The Random Forest (RF) method has been widely used for probability estimation and prediction. RF consists of multiple decision trees, each providing a classification vote for an object. The concept of the random forest was initially proposed by Tin Kam HO of Bell Labs in 1995. It combines random feature selection and bagging techniques to improve performance. Architecture of RF is shown in Figure 1. The random forest algorithm offers several advantages, including:

1) The ensemble learning algorithm using random forest is accurate.
2) Random forest performs well with huge data sets.
3) It can manage a large number of input variables.
4) Random forest calculates the key classification variables.
5) It can deal with missing data.
6) For class unbalanced data sets, Random Forest includes ways for balancing error.
7) With this strategy, generated forests can be preserved for later use.
8) Random forest solves the overfitting issue.
9) RF is less susceptible to outliers in training data.
10) RF allows settings to be simply defined and does away with the requirement for tree pruning.
11) In RF, variable importance and accuracy are automatically generated.

The optimal node to split on is chosen randomly when building individual trees in random forest. Random forest approach is illustrated by the following algorithm.

Algorithm Random Forest

Step 1: pick a fresh bootstrap sample from the training set.

Step 2: On this bootstrap sample, grow on an unpruned tree.

Step 3: Choose the best split at random from among (m try) options at each internal node.

Step 4: If every tree is fully developed, no pruning should be done.

Step 5: As the result of the majority vote cast by all the trees, produce the overall prediction.

$$\hat{y} = arg_k\, max \frac{1}{N} \sum_{n=1}^{N} T_n(x) I(T_n(x) = k)$$

- $\hat{y}$ represents the predicted value or class label.
- $arg_k max$ denotes the argument that maximizes the expression following it.
- 1/N represents the inverse of the number of trees in the random forest (N) and is used to normalize the average.
- $\sum_{n=1}^{N}$ represents the summation over all individual trees in the random forest.
- $T_n(x)$ represents the prediction made by the nth tree in the forest for a given input (x).
- $I(T_n(x) = k)$ is an indicator function that evaluates to 1 if the prediction of the nth tree $T_n(x)$ matches the class label k, and 0 otherwise.

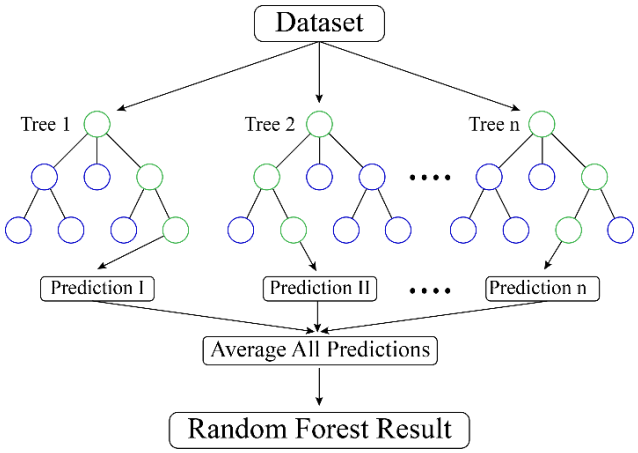

Figure 1. Random Forest Architecture

3.2 XGBoost

XGBoost has become well-known as a potent and effective gradient boosting framework, [26]. Extreme Gradient Boosting, sometimes known as XGBoost, is a technique intended to improve the performance of the gradient boosting approach. The addition of L1 and L2 regularization penalties to the loss function allows for the use of a more regularized model formalization to control overfitting, which is a significant advance. In addition, XGBoost has a "weighted quantile sketch" approach to quickly locate the ideal split points for building trees. Figure 2 depicts the architecture of XGBoost.

Both classification and regression problems frequently use XGBoost, which has the following benefits:

1. High accuracy and strong prediction performance.
2. Scalability and efficiency in handling large datasets with high-dimensional features.
3. Flexibility in objective functions, allowing for a variety of loss functions and evaluation metrics.
4. Support for parallel processing and distributed computing.
5. Built-in regularization to prevent overfitting and improve generalization performance.

The following algorithm provides a high-level overview of how XGBoost works:

Algorithm XGBoost
Step 1: Initialize a set of decision trees with small weights.
Step 2: For each iteration, compute the gradient of the loss function with respect to the current model's output, and use this information to build a new decision tree.
Step 3: Update the weights of the decision trees based on the performance of the new tree.
Step 4: Repeat steps 2 and 3 until the loss function converges or reaches a predetermined threshold.
Step 5: Make predictions using the weighted combination of all the decision trees.

$$\hat{y} = \sum_{k=1}^{K} w_k f_k(x_i)$$

- $\hat{y}$ represents the predicted value or class label.
- $\sum_{k=1}^{k}$ represents the summation over K different components.
- W_k represents the weight or importance assigned to the kth component.
- $f_k(x_i)$ Represents the prediction made by the kth component (also known as a weak learner) for the ith input (x_i)

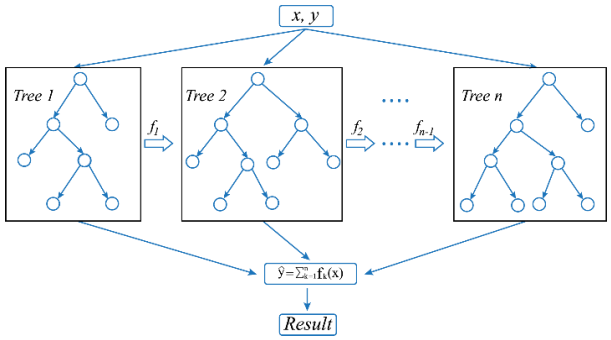

Figure 2. XGBoost Architecture

3.3 Artificial Neural Network

Neural networks, which are nonlinear models of intelligent computational methods, have recently gained recognition as a significant advancement in computing and information processing tools, and their results have shown promise. Feedforward neural networks, which consist of a hidden layer, an appropriate activation function in the hidden layer, and a sufficient number of hidden layer neurons, are a valuable form of artificial neural network capable of estimating any function with arbitrary precision. ANN architecture is shown in Figure 3.

Artificial neural networks typically comprise three different kinds of layers, as follows:

- At the input layer, retrieve the network's raw data [15].
- Hidden layers: The functionality of these layers is influenced by inputs, weights, and the interactions between them. The activation of a hidden unit is determined by the weights between the input and hidden units[15].
- Output layer: output units' performance is influenced by the weight and activity of hidden units as well as the relationship between output and hidden units[15].

$$\hat{y} = f(w\,x\, +\, b)$$

the formula calculates the predicted value or output (ŷ) by passing the weighted sum of the input features (wx) along with the bias term (b) through an activation function (f). The activation function introduces non-linearity into the network and helps in capturing complex patterns and relationships within the data. The weights (w) and biases (b) are learned during the training process of the neural network.

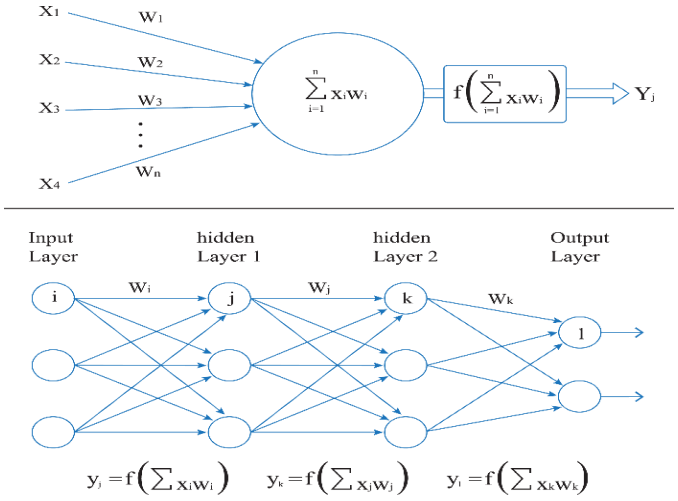

Figure 3. Basic ANN Architecture

3.4 Resnet50

Figure 2 depicts the Residual Network's (ResNet) network architecture. A 50-layer deep CNN called ResNet50 has already been trained on more than a million photos [28]. ResNet has been effectively used for transfer learning in the field of biological image categorization [29].

$$y = F(x, \theta) + x$$

the formula calculates the predicted value or output (y) by applying a function (F) to the input variables (x) with certain parameters (θ), and then adding the input variable (x) to the result. The function (F) and parameters (θ) can vary depending on the specific model or context in which this formula is used.

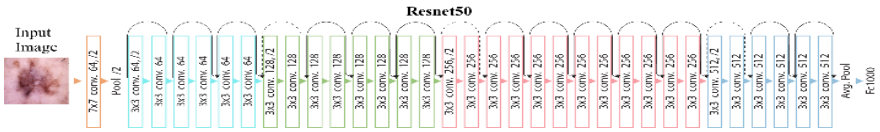

Figure 4. ResNet50 Architecture

3.5 VGG 16

A CNN model named VGG16 was developed using the ImageNet dataset, which contains more than one million images [30]. The 16 deep layer network can classify photos into up to 1000 different categories. The input picture size for this network is 224x224x3, and it provides extensive feature representations for many different types of images [31]. Figure 5 shows the model architecture of VGG 16.

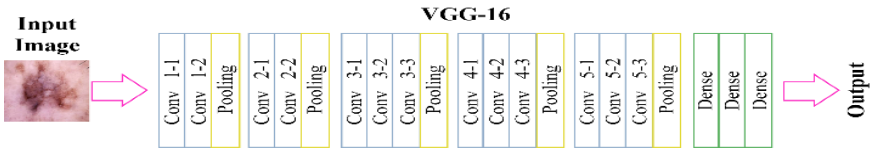

Figure 5. Architecture of VGG16

$$y = f(W_4 f(W_3 f(W_2 f(W_1 x + b_1) + b_2) + b_3) + b_4)$$

- y represents the predicted value or output of the neural network.
- f denotes the activation function applied to the input of each layer.
- w_1, w_2, w_3, w_4 are weight matrices associated with the connections between the layers of the neural network.
- b_1, b_2, b_3, b_4 are bias terms or vectors associated with the respective layers.

4. Proposed Framework

4.1 Existing System

In Numerous predictive systems exist online that can determine the presence of a disease in individuals; however, these systems often exhibit several limitations that must be addressed. Common limitations observed in these systems include a limited number of models, poor user interface quality, and an absence of intuitive user experiences. Moreover, a lack of standardization in data formats and sources, limited generalizability across diverse populations, and a fragmented approach, this fragmentation stems from the fact that certain existing predictive systems specializing in numerical data-based predictions, while others focus exclusively on image-based predictions. This highlights the necessity of a unified platform. It is imperative to overcome these limitations to enhance prediction accuracy, usability, inclusivity, and reliability in disease prediction systems.

4.2 Proposed System

With a comprehensive strategy, our suggested system seeks to solve the shortcomings of current

disease prediction systems. We have carefully selected 8 different diseases, including 5 based on numerical datasets and 3 based on image datasets. For diseases relying on numerical data, we have developed three models: two machine learning models and one neural network. For diseases involving image datasets, we utilize transfer learning techniques with VGG16 and ResNet50 models. These models are deployed on a web-based platform where users can select a specific disease and provide relevant inputs. The system leverages all the trained models to predict the likelihood of the disease. By comparing the predictions of these models, the system generates the most accurate result for the user.

4.3 Methodology

Our dataset consists of both numerical and image data. For numerical data, we used machine learning algorithms such as Random Forest, XGBoost, and Artificial Neural Networks (ANNs). For image data, we used Convolutional Neural Networks (CNNs) with Transfer Learning, using the VGG16 and ResNet50 architectures. The proposed method is shown in Figure 6. In order to deal with different types of data, the proposed methodology is divided into two main categories: one for numerical data, where we applied machine learning and ANNs, and the other for image data, where we will use deep learning with transfer learning via CNNs. We will discuss each of these categories as follows:

```
from keras.preprocessing.image import Image
Data Generator
datagen = Image Data Generator
          (rotation range=20,
          zoom range=0.2,
          horizontal flip=True,
          vertical flip=True,
          width_shift_range=0.1,
          height_shift_range=0.1,
          shear range=0.2,
          fill mode='reflect')
train data = Image Data
Generator.flow_from_directory (
          'train',
          target size=(224, 224),
          batch size=32,
          class mode='categorical')
datagen. Fit(train data)
model. fit generator (datagen. Flow (train data,
batch_size=32),
                steps_per_epoch=Len (train data),
                          epochs=100)
```

Figure 6. Proposed Methodology

Machine Learning with Feedforward Neural Networks (for numerical data) is presented here. Five diseases are based on numerical data which include Heart Attack Analysis, diabetes prediction, stroke prediction, liver and Parkinson disease. The following method is used:

4.3.1 Dataset Description

We used the Kaggle online platform's datasets for this work. Table 1 contains information on the numerical datasets, while Table 2 has information on the image datasets. An Anaconda environment software configuration and a GPU-based system with at least 4 Gb RAM are required for the experimental work.

Table 1. Numerical datasets used for evaluation of models

Dataset	Instances	Features	Classes
Diabetes Dataset	768	8	2
Heart Attack Analysis &	303	13	2

Prediction Dataset			
Indian Liver Patient Records	583	10	2
Parkinson's Disease Data Set	195	23	2
Stroke Prediction Dataset	5110	11	2

Table 2. Image datasets used for evaluation of models

Dataset	No. of Images	Classes
Brain MRI Images for Brain Tumor Detection	253	2
Malaria Cell Images Dataset	27558	2
Melanoma Skin Cancer Dataset of 10000 Images	10000	2

4.3.2 Data Preprocessing

The numerical datasets underwent various preparation stages, including data cleaning (removing duplicates and handling missing values), consistency checks, outlier detection, feature scaling (normalization and standardization), feature selection using Recursive Feature Elimination (RFE) with Random Forest, dealing with imbalanced datasets, and data splitting into training, validation, and test sets.

4.3.3 Feature Selection

For feature driven models, the feature selection is performed using RFE in combination with Random Forest[16].

Random forest is a powerful machine learning algorithm that can perform well for feature selection by evaluating feature importance. However, there are also other methods that can be combined with random forest for feature selection to further improve performance.

One such method is Recursive Feature Elimination (RFE), which can be used in combination with random forest. RFE is an iterative method that removes one or more features at each iteration, trains a model on the remaining features, and evaluates the performance. Until the necessary number of features is attained or performance plateaus, the process is repeated. RFE can be used to pick the most crucial characteristics from a vast set of features, and when combined with random forest, it can assist the feature selection process work better.

Following section presents deep learning with Convolutional Neural Networks and Transfer Learning for image data.

4.3.4 Dataset Description

We used the datasets from the web tool Kaggle in this paper. The datasets are divided training, testing, and validation after being downloaded in their entirety. A GPU-based system with at least 4 Gb RAM and software configured for the Anaconda environment are required for the experimental work.

4.3.5 Data Preprocessing

For each condition, a unique notepad is created. Using Python's PIL package, load datasets and resize the photos to 24x24. To make sure that the image pixel values fell between 0 and 1, each pixel value was normalized by dividing it by 255. In addition, the photos were standardized by taking each pixel's mean pixel value and dividing it by the standard deviation.

4.3.6 Data Augmentation

Data augmentation is a potent method that may be used to expand a dataset and enhance a machine learning model's performance. Our study used the Keras library's Image Data Generator class to perform data augmentation on these three image datasets and applied a set of image transformation techniques to generate new synthetic data points. We then loaded the augmented training data using the flow from directory method and trained our model using a batch size of 32 and a categorical class mode.

5. Experimentation and Results

First, we will present the results for numerical data using machine learning with artificial Neural Networks. The dataset used in this study comprised 80% training data and 20% testing data for Random Forest and XGBoost. For ANN 60% training, 20% validation and 20% testing data. The performance of each algorithm was compared based on accuracy. A summary of the evaluation results of Machine learning models and Feedforward Neural Network is presented in Table 3. As shown, Random Forest has outperformed XGBoost and FNN by generating best accuracy on three out of the five datasets. XGBoost remained second whereas FNN was last in terms of accuracy results.

Table 3: Evaluation of the models based on accuracy using numerical datasets

Disease Model	Random Forest	XGBoost	FNN
Heart Attack	**0.88**	0.81	0.86
Stroke	0.89	**0.91**	0.77
Diabetes	**0.82**	0.78	0.78
Liver Disease	**0.73**	0.70	**0.73**
Parkinson's	0.88	**0.92**	0.88

For image data obtained by employing Deep Learning with Convolutional Neural Networks, the results are presented as follow. For Resnet50 and VGG16, the datasets used included 60% training data, 20% validation data, and 20% testing data. The performance of the models can be found in Table 4. VGG16 has outperformed ResNet50 on all datasets except the Skin Cancer dataset where RestNet50 performance is similar to VGG16.

Table 4: Evaluation of the models based on accuracy using image datasets

Disease Model	ResNet50	VGG16
Brain Tumor	0.91	**0.95**
Malaria	0.88	**0.93**
Skin Cancer	**0.85**	**0.85**

6. Conclusions

This study adopted various efficient machine learning and deep learning models for prediction of multiple common diseases. The early detection of diseases has the potential to significantly enhance life expectancy and mitigate financial burdens. The study used eight different datasets for evaluation of the models. Random Forest, XGBoost and FNN are evaluated on five different numerical datasets whereas VGG16 and ResNet50 are evaluated on three image datasets. Random Orest and VGG16 topped the comparison by generating best results in terms of accuracy on most datasets. In future, we have plans to broaden our analysis to encompass a diverse range of additional illnesses, including lung cancer, breast cancer, bone disease, prostate cancer, colorectal cancer, ovarian cancer, autism spectrum disorder, depression, and anxiety disorders, among others.

References

1. Collaborators, G.B.D., *Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017*. 2018.
2. Rogers, H.W., et al., *Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the US population, 2012*. JAMA dermatology, 2015. **151**(10): p. 1081-1086 %@ 2168-6068.
3. Dorsey, E.R., et al., *Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016*. The Lancet Neurology, 2018. **17**(11): p. 939-953 %@ 1474-4422.
4. Murray, C.J.L., et al., *Global malaria mortality between 1980 and 2010: a systematic analysis*. The Lancet, 2012. **379**(9814): p. 413-431 %@ 0140-6736.
5. Kunjir, A., H. Sawant, and N.F. Shaikh. *Data mining and visualization for prediction of multiple diseases in healthcare*. 2017. IEEE.
6. Saboji, R.G. *A scalable solution for heart disease prediction using classification mining technique*. 2017. IEEE.
7. Kohli, P.S. and S. Arora. *Application of machine learning in disease prediction*. 2018. IEEE.
8. Jena, L. and R. Swain. *Work-in-progress: chronic disease risk prediction using distributed machine learning classifiers*. 2017. IEEE.
9. Chetty, N., K.S. Vaisla, and N. Patil. *An improved method for disease prediction using fuzzy approach*. 2015. IEEE.
10. Ambekar, S. and R. Phalnikar. *Disease risk prediction by using convolutional neural network*. 2018. IEEE.
11. Grampurohit, S., Shalavadi, V., Dhotargavi, V. R., Kudari, M., & Jolad, S. . *Brain tumor detection using deep learning models*. 2020.
12. Shekar, G., Revathy, S., & Goud, E. K. . *Malaria detection using deep learning*. 2020.
13. Ali, M.S., et al., *An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models*. Machine Learning with Applications, 2021. **5**: p. 100036 %@ 2666-8270.
14. Jabbar, M.A., B.L. Deekshatulu, and P. Chandra, *Intelligent heart disease prediction system using random forest and evolutionary approach*. Journal of network and innovative computing, 2016. **4**(2016): p. 175-184 %@ 2160-2174.
15. El_Jerjawi, N.S. and S.S. Abu-Naser, *Diabetes prediction using artificial neural network*. 2018.
16. Genuer, R., J.-M. Poggi, and C. Tuleau-Malot, *Variable selection using random forests*. Pattern recognition letters, 2010. **31**(14): p. 2225-2236 %@ 0167-8655