

Breast Cancer Prediction Using Stacked GRU-LSTM-BRNN

¹Shawni Dutta and ² Prof. Samir Kumar Bandyopadhyay

¹ Lecturer, Department of Computer Science, The Bhawanipur Education Society College, Kolkata, India

²Academic Advisor, The Bhawanipur Education Society College, Kolkata, India

Abstract-

Breast Cancer diagnosis is one of the most studied problems in the medical domain. In the medical domain, cancer diagnosis has been studied extensively which instantiates the need of early prediction of cancer disease. For obtaining advance prediction, health records are exploited and given as input to an automated system. This paper focuses on constructing an automated system by employing deep learning based recurrent neural network models. A stacked GRU-LSTM-BRNN is proposed in this paper that accepts health records of a patient for determining possibility of being affected by breast cancer. Proposed model is compared against other baseline classifiers such as stacked Simple-RNN model, stacked LSTM-RNN model, stacked GRU-RNN model. Comparative results obtained in this study indicate that stacked GRU-LSTM-BRNN yield better classification performance for predictions related to breast cancer disease.

Keywords- Breast Cancer, Predictive Model, Stacked GRU-LSTM-BRNN, LSTM, GRU, RNN.

Introduction-

Adult female population are often affected by Breast cancer (BC) which is one of the more commonly seen cancer types. The population of breast cancer patients includes malignancies of different stages as well as rates of growth. Breast cancer develops from cells lining the milk ducts and slowly grows into a lump or a tumour. Breast cancer may be invasive or non-invasive. Invasive cancer spreads from the milk duct or lobule to other tissues in the breast, whereas, non-invasive ones lack the ability to invade other breast tissues. Non-invasive breast cancer is called "in situ" and may remain inactive for entire lifetime. Briefly BC is often regarded as a heterogeneous disease [1]. Computer aided diagnostic system is often explored and assisted in medical care field. An automated tool is proposed in this paper that assists in clinical care unit by providing early BC prediction. This work is a result of series of discussions with the medical practitioners. Doctors strongly put forward the need of early prediction of breast cancer. Early detection during the onset of the disease can prevent mortality. Early detection can prevent the spreading of the disease thus providing a healthy life to senior citizens.

In India at an average 50% of breast cancer cases are diagnosed at later stages such as III and IV. This diagnosis rate reaches to 12% when it comes to the scenario of developed countries like United States [2]. Another study states that, before the age of 40 years, approximately 7% of women with breast cancer are diagnosed and this disease acquires for more than 40% of all cancer in women in this age group [3]. Knowing the influential factors can increase survival chances among women having breast cancer. This will assist in defining early detection actions, counter measures in health care field. Possible treatment of Breast cancer includes various combinations of chemotherapy, surgery, radiation therapy, hormone therapy and targeted therapy via a multimodality approach.

Hence, detection of this disease at an early stage will assist clinicians in suggesting probable treatments.

For assisting medical fields data mining and knowledge discovery approaches are explored that automatically finds out patterns and relationship among the enormous volume of data. Use of knowledge discovery (KDD) processes is to extract knowledge from data in the context of large databases [4]. Data mining approaches are applicable many areas of medicine, including diagnosis, prognosis and treatment in order to provide benefits. The system is proposed in this paper spontaneously captures previous health records of patient and detects whether the patient can be affected by breast cancer disease or not. Early prediction of this disease is required since cancer is often known as silent killer that develops without any symptoms.

Deep learning techniques are often advantageous because due to self-adaptive structure which is capable to process data with minimal processing. Instead of proceeding feature engineering step manually, this task is assigned to computers which enable non-experts to contribute in the analysis part. Deep learning is an improvement over conventional artificial neural networks since it facilitates the construction of networks by incorporating more than two layers [5]. A deep learning based framework is implemented in this paper that is dedicated for improving the efficiency in breast cancer disease prediction using medical data. A Recurrent Neural Network (RNN) [6] is a type of deep learning model with a feedback loop structure is often helpful in forecasting purpose. A stacked GRU-LSTM model is proposed in this paper that receives past medical records as input and provides prediction regarding the diagnosis of this disease. Gated Recurrent Unit (GRU) [7] and Long-short Term Memory (LSTM) [8] are two variants of RNN which are used in forecasting purpose. The proposed method is evaluated as well as compared with other baseline models such as Simple RNN model, stacked LSTM model, stacked GRU model. Results from proposed model actually provide detection of breast cancer disease at early stage with maximised efficiency. Analysis of the proposed algorithms includes determination of quantitative, qualitative, comparative and complexity measures. The proposed methods have been rigorously tested using dataset.

The paper is motivated strongly for prediction of breast cancer due to:

1. Pre-Processing is done to obtain balanced dataset.
2. GRU and LSTM-BRNN based model is used for higher accuracy.
3. The method is analysed with existing methods to show its performance is superior in all respect than the recent models.

Related Works-

Using neural network and multi-fractal dimension features ultrasound images have been evaluated in order to discriminate benign and malignant breast tumours. This study reported classification result with highest precision of 82.04% [9]. Another study presented the comparative study of clustering methods such as hierarchical clustering, farthest first, LVQ, canopy, and DBSCAN in Weka tool for the diagnosis of breast tumours. Accordingly to the presented result, it is concluded that the farthest first clustering technique had the highest prediction accuracy of 72% [10]. A study proposed deep classification algorithm for mammogram images. An algorithm known as Convolutional Neural Network Improvement for Breast Cancer Classification (CNN-BCC) system has provided using

mammographic images. The algorithm is applied 21 benign, 17 malignant and 183 normal cases provided by Mammographic Image Analysis Society (MIAS) The model achieved 90.50% accuracy [11]. Multiple Instance Learning (MIL) and CNN are combined and presented for BC classification. The experiments were performed on the BreakHis dataset which consists of 8000 microscopic biopsy images of benign and malignant breast tumours. The classification rate was observed as a 92.1% with 40× magnification factor rate [12]. An analysis related to transfer learning has been carried out by employing VGG-16, VGG-19, and ResNet50 deep architectures for the BC histology image classification task. The combination of VGG-16 and logistic regression (LR) yielded the best results with 92.60% accuracy results [13]. Automatic classification of images for breast cancer diagnosis is achieved using Back Propagation Neural Network (BPPN) and radial basis neural networks (RBFN). The accuracies of the BPPN and RBFN are also reported 59.0% and 70.4% respectively [14].

A diagnosis system is proposed for detecting breast cancer by implementing RepTree, RBF Network and Simple Logistic. In test stage, 10-fold cross validation method was applied for evaluating the proposed system performances. The correct classification rate of proposed system is attains 74.5% of efficiency [15]. An extensive study was carried out in by varying the values of k for k- Nearest Neighbour classification technique in order to enhance classification accuracy. Experiments were implemented on breast cancer dataset for early disease detection [16]. Delen et al. investigated the use of artificial neural networks, decision trees and logistic regression to develop prediction models for breast cancer survival. 10-fold cross-validation methods are explained to measure the unbiased estimate of the three prediction models for performance comparison purposes. The results indicated that the decision tree turns out to be the best predictor with 93.6% accuracy [17]. Another study investigated use of three algorithms like Decision Tree (C4.5), Artificial Neural Networks (ANN), and Support Vector Machine (SVM) are implemented in order to find classification accuracy in breast cancer dataset. Comparative study analyses that SVM produces higher accuracy in classification [18].

As mentioned in the related works, several researches have been carried out to improve the performance of computational approaches to diagnosing BC and to ensure the development of a of diagnosis system. To accompany this diagnosis system, a novel system is proposed in this paper. By incorporating GRU as well as LSTM layers are assembled within a single platform in order to provide prediction of breast cancer in advance.

Background-

Deep learning provides a multi-layered hierarchical data representation typically in the form of a neural network by assembling more than two layers. Neural network model is built by coalescing multiple layers with linear or non-linear activation functions those are trained together for achieving complex problem solving approach. Activation functions are capable of executing diverse computations and produce outputs within a definite range. In other words, activation function is a step that maps input signal into output signal [19]. Sigmoid and ReLu are two popular activation functions that are described as follows-

- Sigmoid activation function [19] transforms input data in the range of 0 to 1 and it is shown in equation (1).

$$f(x) = 1/(1 + \exp^{-x}) \quad (1)$$

- ReLu activation function [19] is a faster learning Activation function which is the most successful and widely used function. It performs a threshold operation to each input element where values less than zero are set to zero whereas the values greater or equal to zeros kept as intact and it is shown in equation (2).

$$f(x) = \max(0, x) = \begin{cases} X_i, & \text{if } X_i \geq 0 \\ 0, & \text{if } X_i < 1 \end{cases} \quad (2)$$

Recurrent neural network (RNN) is a type of neural network architecture that processes both sequential and parallel information. Similar operations like human brain can be simulated by incorporating memory cells to the neural network. Another RNN called Bidirectional RNN (BRNN) are designed to access input sequences whose starts and ends are known in advance. Because RNN can only take information from the previous context, further improvements can be made using Bi-RNN. The Bi-RNN can handle two sources of information. While considering both past and future context of each sequence element into justification, one RNN processes the sequence from start to end, the other backwards from end to start [20].

There are alternatives from RNN depending on the gating units, such as LSTM-RNN and GRU-RNN.

Long short-term memory (LSTM) neural network is a kind of RNN that implements context based prediction which is not considered in traditional RNN. In other words, LSTM is capable to eliminate the problem of vanishing gradient by training RNN. LSTM has a good potential to regulate gradient flow as well as better preservation of long-range dependencies. Every cell in LSTM is comprised of gates that determine when to remember input, when to remember or forget the value and when it should output the value. Depending on the performance, there are variants in gates such as input gate, output gate and forget gate. The input gate blocks a value from entering into next layer when a value close to zero is generated by this gate. This input gate simply eliminates the value from the net input. Forget gate remembers value until greater value than zero is generated by forget gate. The block effectively forgets the value what it was remembering when close-to-zero value is produced. The output gate determines when the unit should output the value in its memory [8][21].

Given $x_t = (x_1, \dots, x_T)$ be an input sequence, output sequence $y_t = (y_1, \dots, y_T)$ can be obtained through the following equations (3-8) where W is the weight matrices, b is the bias vector, σ states the sigmoid function, and i, f, o, c are the input gate, forget gate, output gate, and cell activation vectors, respectively. W_{ix} denotes the weight's matrix from the input gate to the input. W_{ic}, W_{fc}, W_{oc} denote diagonal weight matrices for peephole connections, In addition, \odot and \emptyset indicate element-wise multiplication and softmax activation function for the LSTM.

$$i_t = \sigma [(W_{ix} \times x_t) + (W_{ir} \times r_{t-1}) + (W_{ic} \times c_{t-1}) + b_i] \quad (3)$$

$$f_t = \sigma [(W_{fx} \times x_t) + (W_{fr} \times r_{t-1}) + (W_{fc} \times c_{t-1}) + b_f] \quad (4)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh [(W_{cx} \times x_t) + (W_{cr} \times r_{t-1}) + b_c] \quad (5)$$

$$o_t = \sigma [(W_{ox} \times x_t) + (W_{or} \times r_{t-1}) + (W_{oc} \times c_t) + b_o] \quad (6)$$

$$r_t = o_t \odot \tanh (c_t) \quad (7)$$

$$y_t = \emptyset(W_{yr} \times r_t + b_y) \quad (8)$$

Gated Recurrent Unit (GRU) is quite similar to LSTM, where the gating units of GRU control the flow of information inside the unit, without considering separate memory cells. Like LSTM, GRU lacks of having memory cells in it and it has a lesser number of gates which are activated using current input as well as previous output. GRU controls the information flow from the previous activation while computing the new, candidate activation, but does not independently control the amount of the candidate activation being added. As compared to LSTM, GRU has better convergence rate due to reduction of parameters and in some cases GRU outperforms well over LSTM models [7].

While designing the Deep models, including Dropout layers are often useful in order to reduce over-fitting problem. Dropout layers randomly deactivate a fraction of the units or connections in a network during each of the training iterations [5]. After configuring neural network models, training process is executed. The training process goes through one cycle known as an epoch where the dataset is partitioned into smaller sections. An iterative process is executed through a couple of batch size that considers subsections of training dataset for completing epoch execution [22]. Since this entire framework is inclined towards solving binary classification problem, binary cross entropy function is used as training criterion. Binary cross entropy measures the distance from the true value (which is either 0 or 1) to the prediction for each of the classes and then averages these class-wise errors to obtain the final loss [23].

While stacking RNN based layers into a single framework, employing an optimizer is necessary. Adam is one of the popular optimizers that is computationally efficient with lower memory requirement and also easy to implement. This algorithm is applicable for first-order gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order moments. This algorithm is quite well accepted due to its applicability on non-stationary objectives and problems with very noisy and/or sparse gradients [24].

Dataset Used-

In this context, Breast Cancer Wisconsin (Diagnostic) Data Set is collected from UCI [25]. The dataset consists of 1470 number of sample records and each of which can be framed as collection of attributes that include several criteria for detecting patients having breast cancer symptoms. The dataset can be formulated as collection of attributes that include several criteria for detecting breast cancer tendency such as radius (mean of distances from centre to points on the perimeter), texture (standard deviation of gray-scale values), perimeter, area, smoothness (local variation in radius lengths), compactness ($\text{perimeter}^2 / \text{area} - 1.0$), concavity (severity of concave portions of the contour), concave points (number of concave portions of the contour), symmetry, fractal dimension ("coastline approximation" - 1), diagnosis. However, the attribute 'diagnosis' is utilized as output class of the prediction which contains the class either benign or malignant. Following diagram Fig.1 shows overall understanding of the dataset.

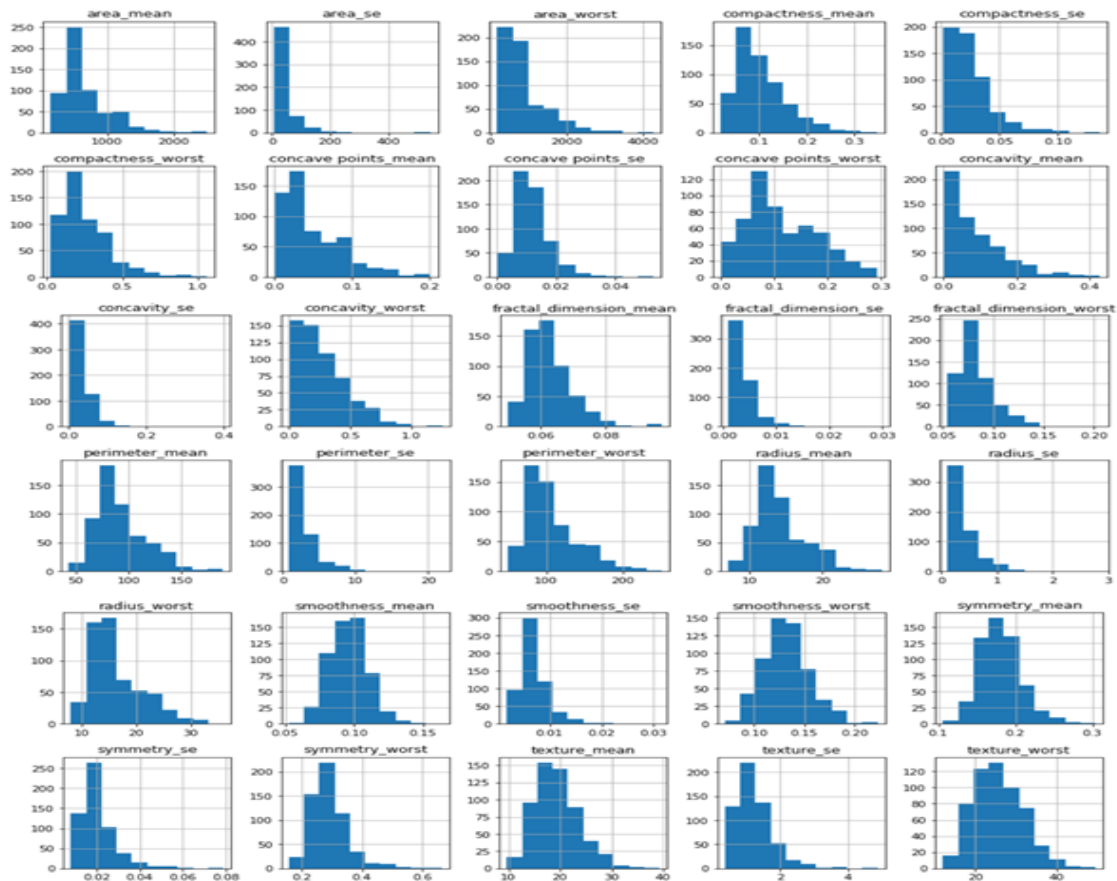


Figure1. Visual Representation of Dataset Attributes.

Proposed Methodology-

Data mining techniques are applied in this paper for the purpose of breast cancer prediction. Two major processes such as data pre-processing and classification/clustering are essential steps in data mining process. Any classification or clustering phase is followed by implementing data pre-processing where redundant or irrelevant information are eliminated from the original. Classification or clustering step are executed in order to obtain the task of prediction, estimation, etc. Following section explains application of data mining techniques for breast cancer prediction.

1. Dataset Pre-processing-

Once data are collected, pre-processing techniques are applied in order to obtain balanced dataset. Pre-processing techniques include checking and handling missing value, scaling some attributes. Existing 'nan' values are also handled for this data. Some of the attributes such as 'id', 'Unnamed :32' are eliminated since they have no contribution in prediction. Applying these pre-processing techniques will yield a transformed dataset that can be fitted to classifier. Next, feature scaling of relevant attributes are performed which enhances the efficiency while fitting to a classifier. The transformed dataset is partitioned into training set and testing dataset which is obtained by partitioning the transformed dataset with the ratio of 7:3. Training data is fitted to RNN models and predictions are received for the test dataset.

2. Methodology and Implementation-

The main objective of proposed classifier is to predict whether a patient has breast cancer disease or not using deep learning techniques. The classification process aims to detect benign or malignant cancer patients. The proposed method implements GRU and LSTM-BRNN based framework for such predictions.

A stacked GRU-LSTM based model is proposed in this paper that contains alternate sequence of GRU and LSTM layers along with four dense layers. The LSTM and GRU layers are implemented as bidirectional RNN. Except the dense layers, LSTM and GRU layers are followed by dropout layers. Use of dropout layers prevents this model from over-fitting.

While designing this model it is necessary to tune hyper-parameters in order to achieve maximized efficiency. This section describes specification of the model along with its hyper-parameters. The proposed model consists of having four LSTM and GRU layers with 128,64,32,16 numbers of units respectively. Each of these layers is followed by layer having dropout rate of 20%. Next, four dense layers are stacked in this model with 8,4,2,1 number of nodes respectively. Except the dropout layers, sigmoid activation function is used in the layers. Finally these aforementioned layers are compiled using adam optimiser through 100 epochs and with a batch size of 64. Adjustment of the hyper-parameters assists the model to attain best predictive result. The neural network receives a total of 305,785 numbers of parameters and trains those parameters in order to obtain prediction. The detailed description of the model is given in Figure 2.

Layer (type)	Output Shape	Param #
bidirectional_5 (Bidirection)	(None, 30, 256)	99840
dropout_5 (Dropout)	(None, 30, 256)	0
bidirectional_6 (Bidirection)	(None, 30, 128)	164352
dropout_6 (Dropout)	(None, 30, 128)	0
bidirectional_7 (Bidirection)	(None, 30, 64)	30912
dropout_7 (Dropout)	(None, 30, 64)	0
bidirectional_8 (Bidirection)	(None, 32)	10368
dropout_8 (Dropout)	(None, 32)	0
dense_5 (Dense)	(None, 8)	264
dense_6 (Dense)	(None, 4)	36
dense_7 (Dense)	(None, 2)	10
dense_8 (Dense)	(None, 1)	3

Figure2. Description of Stacked GRU-LSTM-BRNN model

3. Baseline classifiers and Implementation-

The proposed model is evaluated against set of other RNN models such as Simple RNN, Stacked LSTM model, and Stacked GRU model. All the models are provided with its corresponding description and implementation details.

a. Simple Recurrent Neural Network-

This model is designed by stacking four simple RNN layers into a single platform. Dropout layers are used after each RNN layers with a rate of 20%. However, these layers are followed by incorporating four dense layers. First four RNN layers and last dense layer receives 'sigmoid' as activation function. A total of 33,065 parameters are fed into this model during training phase. Summarised description of this model is given as Figure3.

Layer (type)	Output Shape	Param #
simple_rnn_1 (SimpleRNN)	(None, 30, 128)	16640
dropout_1 (Dropout)	(None, 30, 128)	0
simple_rnn_2 (SimpleRNN)	(None, 30, 64)	12352
dropout_2 (Dropout)	(None, 30, 64)	0
simple_rnn_3 (SimpleRNN)	(None, 30, 32)	3104
dropout_3 (Dropout)	(None, 30, 32)	0
simple_rnn_4 (SimpleRNN)	(None, 16)	784
dropout_4 (Dropout)	(None, 16)	0
dense_1 (Dense)	(None, 8)	136
dense_2 (Dense)	(None, 4)	36
dense_3 (Dense)	(None, 2)	10
dense_4 (Dense)	(None, 1)	3

Figure3. Description of Stacked Simple-RNN model

b. Stacked LSTM-RNN model-

A stacked LSTM model is implemented that contains four LSTM layers, each of them are followed by dropout layers. Again four dense layers are also incorporated into the model. LSTM layers and the last dense layer is implemented using 'sigmoid' activation function. Once this model is implemented by choosing appropriate hyper-parameters, it receives 131,705 parameters and trains those parameters for acquiring prediction result. This model is concise in figure4.

Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(None, 30, 128)	66560
dropout_1 (Dropout)	(None, 30, 128)	0
lstm_2 (LSTM)	(None, 30, 64)	49408
dropout_2 (Dropout)	(None, 30, 64)	0
lstm_3 (LSTM)	(None, 30, 32)	12416
dropout_3 (Dropout)	(None, 30, 32)	0
lstm_4 (LSTM)	(None, 16)	3136
dropout_4 (Dropout)	(None, 16)	0
dense_1 (Dense)	(None, 8)	136
dense_2 (Dense)	(None, 4)	36
dense_3 (Dense)	(None, 2)	10
dense_4 (Dense)	(None, 1)	3

Figure4. Description of Stacked LSTM-RNN model

c. Stacked GRU-RNN model-

Like the other two baseline classifiers, this model also consists of sequence of GRU as well dropout layers. A series of four dense layers are also stacked into this model. The GRU layers and the final output layer accept 'sigmoid' as activation function just like previous two models. Designing the model is capable to draw prediction by considering 98,825 parameters during training phase. Brief description of the model is shown in figure5.

Layer (type)	Output Shape	Param #
gru_1 (GRU)	(None, 30, 128)	49920
dropout_1 (Dropout)	(None, 30, 128)	0
gru_2 (GRU)	(None, 30, 64)	37056
dropout_2 (Dropout)	(None, 30, 64)	0
gru_3 (GRU)	(None, 30, 32)	9312
dropout_3 (Dropout)	(None, 30, 32)	0
gru_4 (GRU)	(None, 16)	2352
dropout_4 (Dropout)	(None, 16)	0
dense_1 (Dense)	(None, 8)	136
dense_2 (Dense)	(None, 4)	36
dense_3 (Dense)	(None, 2)	10
dense_4 (Dense)	(None, 1)	3

Figure5. Description of Stacked GRU-RNN model

4. General Model Structure-

This section provides the general structure of RNN model architecture. All the aforementioned models are implemented with same number of layers, along with same number of units. Number of epochs and batch size are kept as fixed for all these models. The essential part is the RNN layer type that gets differ for each model. The implemented models are following same structure in terms of number of nodes present in each layers, present dropout layer rate, epochs, batch size, and optimizer. This will provide similar platform to compare models in terms of the prediction. The specified parameters common to each of the implemented models are defined in Table1.

Layer Numbers->	1	2	3	4	5	6	7	8
Number of Nodes	128	64	32	16	8	4	2	1
Type of Layer	RNN Layer				Dense Layers			
Dropout Layer Present?	Yes [Dropout Rate- 20%]				No			
Batch Size	64							
Epochs	100							
Loss	Binary Cross-entropy							
Optimizer	Adam Optimizer							

Table1. General Structure Implemented for RNN-based Model

5. Performance Measure Metrics-

While evaluating performance skill of a model, it is necessary to employ some metrics to justify the evaluation. Use of these metrics will assist in identifying best problem solving approach. The metrics are explained as follows-

1. Accuracy-

Accuracy is a metric that ascertains the ratio of true predictions over the total number of instances considered. However, the accuracy may not be enough metric for evaluating model's performance since it does not consider wrong predicted cases with different weights [26].

2. F1-Score-

For compensating the above mentioned problem, we consider two more metrics known as, Recall and Precision. Precision identifies the ratio of correct positive results over the number of positive results predicted by the classifier. Recall denotes the number of correct positive results divided by the number of all relevant samples. *F1-Score or F-measure* is a parameter that is concerned for both recall and precision and it is calculated as the harmonic mean of precision and recall [26].

3. Cohen-kappa Score-

Cohen-Kappa Score is also taken into consideration as an evaluating metric in this paper. This metric is a statistical measure that finds out inter-rater agreement for qualitative items for classification problem [27].

4. MSE-

Mean Squared Error (MSE) is another evaluating metric that measures absolute differences between the prediction and actual observation of the test samples [26].

Mathematically, the aforementioned metrics can be defined as follows with given True Positive, True Negative, False Positive, False Negative as TP, TN, FP, FN respectively-

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+TP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{F1- Measure or F1-Score} = \frac{2 * \text{Recall} * \text{Precision}}{(\text{Recall} + \text{Precision})}$$

$$\text{Cohen-Kappa Score} = \frac{(p_o - p_e)}{(1 - p_e)}$$

where p_o denotes relative observed agreement among raters and p_e is the probability of agreement by chance.

$$\text{MSE} = \left(\sum_{i=1}^N (X_i - X_i')^2 / N \right) \text{ where } X_i \text{ is the actual value and } X_i' \text{ is the predicted value.}$$

Lower value of MSE and higher values of accuracy, F1-Score, and Cohen-kappa score signifies a better performing model.

Experimental Results-

During the training process, RNN models are learned by feeding training dataset. Training process is evaluated in terms of loss and accuracy over each epoch. Proposed model along with other baseline classifiers are trained through 100 epochs. Accuracy and loss acquired through the epochs for each of the models are indicated in figure 5 to 8.

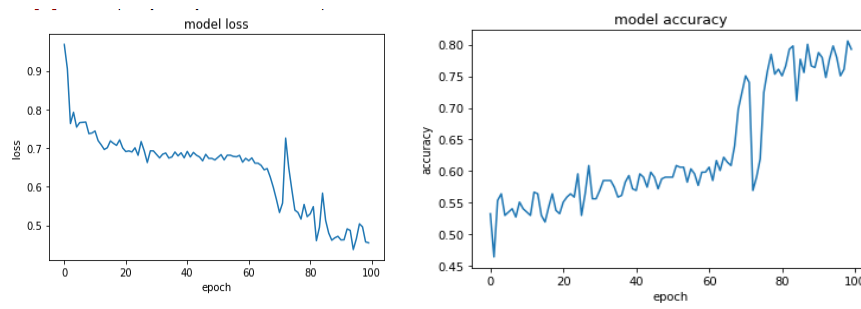


Figure5. Stacked Simple-RNN Model Loss and Accuracy plot obtained over epochs.

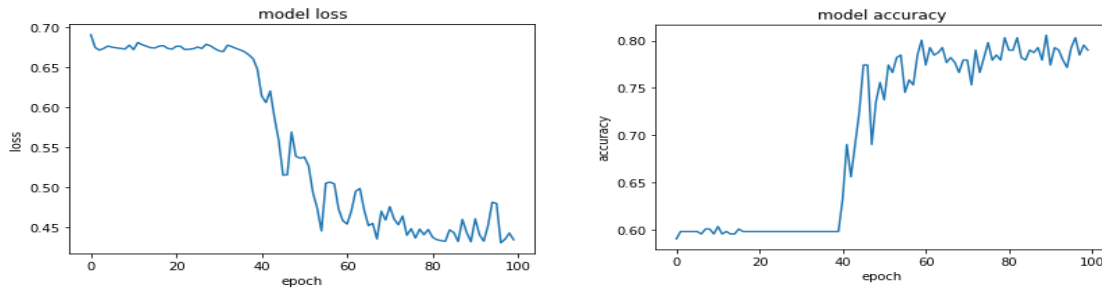


Figure6. Stacked LSTM Model Loss and Accuracy plot obtained over epochs.

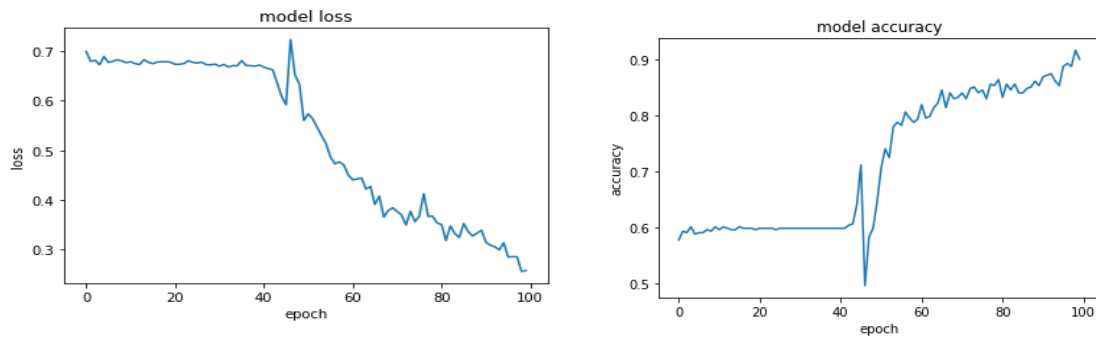


Figure7. Stacked GRU Model Loss and Accuracy plot obtained over epochs.

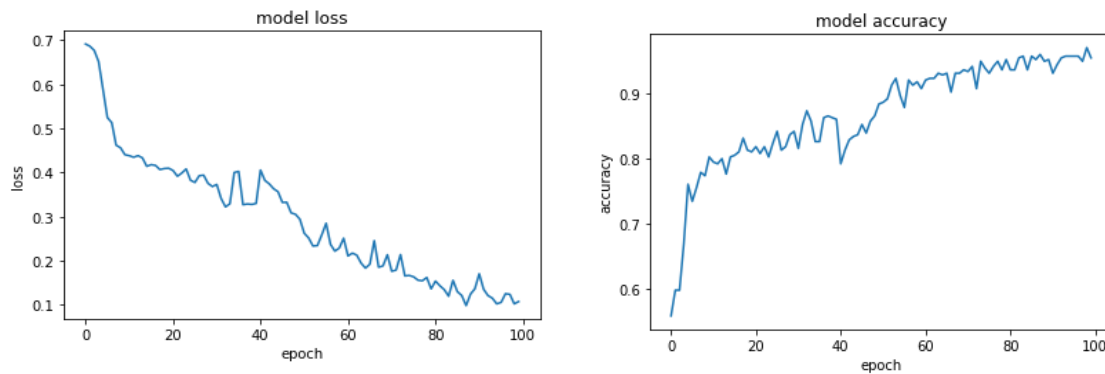


Figure8. Proposed Stacked GRU-LSTM Model Loss and Accuracy plot obtained over epochs.

The proposed stacked GRU-LSTM BRNN model is implemented and evaluated in terms of accuracy, F1-Score, MSE and Cohen-Kappa Score. Loss occurred during testing is also measured. This model is later compared with other baseline classifiers known as Simple RNN based model, Stacked LSTM model, and Stacked GRU model. All the implemented models are demonstrated in terms of performance using evaluating metrics after completion of 100 epochs of training. After this training session using training data, predictions are obtained for the test dataset. These predictions are compared with actual observed data which instantiates the evaluation of deep models with respect to employed metrics. The comparative study is shown in table2. From the comparative study it is clear that the proposed model indicates much promising result over other classifiers.

Evaluating Metrics	Simple RNN Model	Stacked LSTM Model	Stacked GRU Model	Proposed Stacked GRU-LSTM Model
Test Loss	0.458601399305019	0.3743316245839951	0.21349462422918766	0.09276566210579365
Accuracy	77.13%	84.57%	93.62%	97.34%
F1-Score	0.77	0.85	0.94	0.97
MSE	0.23	0.15	0.06	0.03
Cohen-Kappa Score	0.54	0.65	0.85	0.94

Table2. Performance Summarization of All RNN- based models.

As shown in table 2, Stacked LSTM deep model provides better classification result than stacked simple RNN deep model because the LSTM is superior over simple RNN due to its structure in long-term dependencies. Again it is already established, GRU has better performance over LSTM. Therefore an improvement of classification result is observed in Stacked GRU deep model. By incorporating the advantages of LSTM as well as GRU into a single platform, even more improvised result is obtained in terms of classification.

Conclusions-

Breast cancer is severe disease that needs to be handled carefully. Detection of this disease at an early stage is quite helpful in saving patients life. The objective of this study is to detect the feasibility of utilising previous medical records and determine the probability of being affected by cardiac arrest. Using deep learning methods, a stacked GRU-LSTM layer based model is proposed and implemented in this paper. Interfering attributes those have impact on this disease are considered while designing the model with necessary parameter tuning. The highest value of accuracy, f1-score, cohen-kappa score and lowest value of test loss and MSE is attained by the stacked GRU-LSTM model which denotes superiority of the model over other baseline classifiers. The proposed method achieves promising result with an accuracy of 97.34%, F1-Score of 0.97, Cohen-Kappa Score of 0.94 and MSE of 0.03.

References-

- [1] M. Sibbering and C. A. Courtney, "Management of breast cancer: Basic principles," *Surg. (United Kingdom)*, vol. 34, no. 1, pp. 25–31, 2016, doi: 10.1016/j.mpsur.2015.10.005.
- [2] J. Sathwara, S. Bobdey, and G. B., "Breast cancer survival studies in India: a review," *Int. J. Res. Med. Sci.*, vol. 4, no. 8, pp. 3102–3108, 2016, doi: 10.18203/2320-6012.ijrms20162266.

- [3] C. K. Anders, R. Johnson, J. Litton, M. Phillips, and A. Bleyer, "Breast Cancer Before Age 40 Years," *Semin. Oncol.*, vol. 36, no. 3, pp. 237–249, 2009, doi: 10.1053/j.seminoncol.2009.03.001.
- [4] scar Marbn, G. Mariscal, and J. Segovi, "A Data Mining & Knowledge Discovery Process Model," *Data Min. Knowl. Discov. Real Life Appl.*, no. May 2014, 2009, doi: 10.5772/6438.
- [5] S. H.-I. Shen Dinggang, Wu Gurrong, "Deep Learning in Medical Image Analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, no. March, pp. 221–248, 2017, doi: 10.1146/annurev-bioeng-071516.
- [6] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," *Phys. D Nonlinear Phenom.*, vol. 404, no. March, pp. 1–43, 2020, doi: 10.1016/j.physd.2019.132306.
- [7] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," pp. 1–9, 2014.
- [8] D. Soutner and L. Müller, "Application of LSTM neural networks in language modelling," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8082 LNAI, pp. 105–112, 2013, doi: 10.1007/978-3-642-40585-3_14.
- [9] M. A. Mohammed, B. Al-Khateeb, A. N. Rashid, D. A. Ibrahim, M. K. Abd Ghani, and S. A. Mostafa, "Neural network and multi-fractal dimension features for breast cancer classification from ultrasound images," *Comput. Electr. Eng.*, vol. 70, pp. 871–882, 2018, doi: 10.1016/j.compeleceng.2018.01.033.
- [10] R. Delshi Howsalya Devi and P. Deepika, "Performance comparison of various clustering techniques for diagnosis of breast cancer," *2015 IEEE Int. Conf. Comput. Intell. Comput. Res. ICCIC 2015*, 2016, doi: 10.1109/ICCIC.2015.7435711.
- [11] F. F. Ting, Y. J. Tan, and K. S. Sim, "Convolutional neural network improvement for breast cancer classification," *Expert Syst. Appl.*, vol. 120, pp. 103–115, 2019, doi: 10.1016/j.eswa.2018.11.008.
- [12] P. J. Sudharshan, C. Petitjean, F. Spanhol, L. E. Oliveira, L. Heutte, and P. Honeine, "Multiple instance learning for histopathological breast cancer image classification," *Expert Syst. Appl.*, vol. 117, pp. 103–111, 2019, doi: 10.1016/j.eswa.2018.09.049.
- [13] Shallu and R. Mehra, "Breast cancer histology images classification: Training from scratch or transfer learning?," *ICT Express*, vol. 4, no. 4, pp. 247–254, 2018, doi: 10.1016/j.icte.2018.10.007.
- [14] S. Kaymak, A. Helwan, and D. Uzun, "Breast cancer image classification using artificial neural networks," *Procedia Comput. Sci.*, vol. 120, pp. 126–131, 2017, doi: 10.1016/j.procs.2017.11.219.
- [15] V. Chaurasia and S. Pal, "International Journal of Computer Science and Mobile Computing Data Mining Techniques: To Predict and Resolve Breast Cancer Survivability," *Int. J. Comput. Sci. Mob. Comput.*, vol. 3, no. 1, pp. 10–22, 2014.
- [16] S. AhmedMedjahed, T. Ait Saadi, and A. Benyettou, "Breast Cancer Diagnosis by using k-Nearest Neighbor with Different Distances and Classification Rules," *Int. J. Comput. Appl.*, vol. 62, no. 1, pp. 1–5, 2013, doi: 10.5120/10041-4635.
- [17] D. Delen, G. Walker, and A. Kadam, "Predicting breast cancer survivability: A comparison of

- three data mining methods,” *Artif. Intell. Med.*, vol. 34, no. 2, pp. 113–127, 2005, doi: 10.1016/j.artmed.2004.07.002.
- [18] A. LG and E. AT, “Using Three Machine Learning Techniques for Predicting Breast Cancer Recurrence,” *J. Heal. Med. Informatics*, vol. 04, no. 02, 2013, doi: 10.4172/2157-7420.1000124.
- [19] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, “Activation Functions: Comparison of trends in Practice and Research for Deep Learning,” pp. 1–20, 2018.
- [20] Z. C. Lipton, J. Berkowitz, and C. Elkan, “A Critical Review of Recurrent Neural Networks for Sequence Learning,” pp. 1–38, 2015.
- [21] Ü. Budak, Z. Cömert, Z. N. Rashid, A. Şengür, and M. Çıbuk, “Computer-aided diagnosis system combining FCN and Bi-LSTM model for efficient breast cancer detection from histopathological images,” *Appl. Soft Comput. J.*, vol. 85, no. September, p. 105765, 2019, doi: 10.1016/j.asoc.2019.105765.
- [22] Y. You, J. Hseu, C. Ying, J. Demmel, K. Keutzer, and C. J. Hsieh, “Large-batch training for LSTM and beyond,” *Int. Conf. High Perform. Comput. Networking, Storage Anal. SC*, pp. 1–15, 2019, doi: 10.1145/3295500.3356137.
- [23] K. Janocha and W. M. Czarnecki, “On loss functions for deep neural networks in classification,” *Schedae Informaticae*, vol. 25, pp. 49–59, 2016, doi: 10.4467/20838476SI.16.004.6185.
- [24] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, 2015.
- [25] Michael Kahn, St. Louis. UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science.
- [26] H. M and S. M.N, “A Review on Evaluation Metrics for Data Classification Evaluations,” *Int. J. Data Min. Knowl. Manag. Process*, vol. 5, no. 2, pp. 01–11, 2015, doi: 10.5121/ijdkp.2015.5201.
- [27] S. M. Vieira, U. Kaymak, and J. M. C. Sousa, “Cohen’s kappa coefficient as a performance measure for feature selection,” *2010 IEEE World Congr. Comput. Intell. WCCI 2010*, no. May 2016, 2010, doi: 10.1109/FUZZY.2010.5584447.