

Breast Cancer Prediction Using Stacked GRU-LSTM-BRNN

Shawni Dutta¹, Jyotsna Kumar Mandal², Tai Hoon Kim³, Samir Kumar Bandyopadhyay^{4*}

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

²Department of Computer Science & Engineering, University of Kalyani, Kalyani, West Bengal, India

³School of Economics and Management, Beijing Jiaotong University, China

⁴Department of Computer Science, The Bhawanipur Education Society College, Kolkata, India

Abstract – Breast Cancer diagnosis is one of the most studied problems in the medical domain. Cancer diagnosis has been studied extensively, which instantiates the need for early prediction of cancer disease. To obtain advance prediction, health records are exploited and given as input to an automated system. The 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 the possibility of being affected by breast cancer. The 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 the stacked GRU-LSTM-BRNN model yields better classification performance for predictions related to breast cancer disease.

Keywords – Breast cancer, GRU, LSTM, predictive model, RNN, stacked GRU-LSTM-BRNN.

I. INTRODUCTION

Adult female population is 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 [1]. In India, an average 50 % of breast cancer cases are diagnosed at later stages such as III and IV. This diagnosis rate reaches 12 % when it comes to the scenario of developed countries, such as the 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 more than 40 % of all cancer in women in this age group [3]. Computer aided diagnostic systems are often explored and assisted in the medical care field. An automated tool is proposed in this paper that assists in the clinical care unit by providing early BC prediction. This study is a result of a series of discussions with the medical practitioners. Doctors strongly put forward the need for early prediction of breast cancer. Early detection during the onset of the disease can prevent mortality. Knowing the influential factors can increase survival chances among women having breast cancer. This will

assist in defining early detection actions, counter measures in the healthcare 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.

To assist the medical field, data mining and knowledge discovery approaches are explored that automatically find patterns and relationships 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 in many areas of medicine, including diagnosis, prognosis and treatment in order to provide benefits. The system proposed in this paper spontaneously captures previous health records of patients 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 a silent killer that develops without any symptoms.

Machine Learning (ML) techniques simulate the process of data mining by teaching the computer how to comprehend a complex problem. Deep learning techniques, a subset of ML, are often advantageous because of the self-adaptive structure, which is capable of processing data with minimal processing. Instead of proceeding with the feature engineering step manually, this task is assigned to computers, which enable non-experts to contribute to the analysis part. Deep learning is 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 to 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 that is often helpful in forecasting purposes. 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

* Corresponding author's e-mail: 1954samir@gmail.com

for forecasting purposes. 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 of the proposed model actually provide the detection of breast cancer disease at an 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 a dataset.

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

- 1) pre-processing is done to obtain a balanced dataset;
- 2) GRU and LSTM-BRNN based models are used for higher accuracy;
- 3) the method is analysed with existing methods to show that its performance is superior in all respects than the recent models.

II. RELATED WORKS

Neural networks and ultrasound images of multi-fractal dimension features were evaluated in order to discriminate between benign and malignant breast tumours. This study reported classification results with the highest precision of 82.04 % [9]. In [10], the comparative study of clustering methods was performed, such as hierarchical clustering, farthest first, LVQ, canopy, and DBSCAN in Weka tool for the diagnosis of breast tumours. According to the presented result, it was concluded that the farthest first clustering technique had the highest prediction accuracy of 72 % [10]. A study proposed a deep classification algorithm for mammogram images. An algorithm known as Convolutional Neural Network Improvement for Breast Cancer Classification (CNN-BCC) system was provided using mammographic images. The algorithm was applied in 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 were combined and presented for BC classification. The experiments were performed on the BreakHis dataset, which consisted of 8000 microscopic biopsy images of benign and malignant breast tumours. The classification rate was observed as a 92.1 % with 40 \times magnification factor rate [12]. An analysis related to transfer learning was 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 [13]. Automatic classification of images for breast cancer diagnosis was achieved using a Back Propagation Neural Network (BPNN) and Radial Basis Neural Networks (RBFNs). The accuracies of the BPNN and RBFN were also reported 59.0 % and 70.4 %, respectively [14].

A diagnosis system was proposed for detecting breast cancer by implementing RepTree, RBF Network and Simple Logistic. During the test stage, a 10-fold cross validation method was applied for evaluating the proposed system performance. The correct classification rate of the proposed system attained 74.5 % of efficiency [15]. An extensive study was carried out

by varying the values of k for k -Nearest Neighbour classification technique in order to enhance classification accuracy. Experiments were implemented on a 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 were explained to measure the unbiased estimate of the three prediction models for performance comparison purposes. The results indicated that the decision tree turned out to be the best predictor with 93.6 % accuracy [17]. Another study investigated the use of three algorithms such as Decision Tree (C4.5), Artificial Neural Networks (ANN), and Support Vector Machine (SVM) in order to find classification accuracy in a breast cancer dataset. The comparative study analysed that SVM produced higher accuracy in classification [18].

As mentioned in the related works, several studies were carried out to improve the performance of computational approaches to diagnosing BC and to ensure the development of a 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.

III. 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 models are built by coalescing multiple layers with linear or non-linear activation functions that are trained together for achieving complex problem solving approaches. Activation functions are capable of executing diverse computations and produce outputs within a definite range. In other words, an activation function is a step that maps an input signal into an output signal [19]. Sigmoid and ReLu are two popular activation functions.

Recurrent Neural Network (RNN) is a type of neural network architecture that processes both sequential and parallel information. Similar operations like the human brain can be simulated by incorporating memory cells to the neural network. Another RNN called Bidirectional RNN (BRNN) is designed to access input sequences whose starts and ends are known in advance. Since 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 of eliminating the problem of vanishing gradients by training RNN. LSTM has a good potential to regulate a gradient flow as well as to ensure a better preservation of long-range dependencies. Every cell in LSTM consists 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 the 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 the value until a greater value than zero is generated by the forget gate. The block effectively forgets the value that it has been remembering when a close-to-zero value is produced. The output gate determines when the unit should output the value in its memory [8], [21].

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 memory cells in it and it has a fewer number of gates which are activated using the current input as well as the 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 a better convergence rate due to reduction of parameters and in some cases GRU outperforms well over LSTM models [7].

Designing the deep models, including dropout layers, is often useful in order to reduce an 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, the 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 sizes that consider subsections of a training dataset for completing epoch execution [22]. Since this entire framework is inclined towards solving a binary classification problem, binary cross-entropy function is used as a 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, it is necessary to employ an optimizer. Adam is one of the popular optimizers that is computationally efficient with lower memory requirements 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 to non-stationary objectives and problems with very noisy and/or sparse gradients [24].

IV. DATASET USED

In this context, Breast Cancer Wisconsin (Diagnostic) Dataset is collected from UCI [25]. The dataset consists of 569 sample records and each can be framed as a collection of attributes that include several criteria for detecting patients having breast cancer symptoms. The dataset consists of an attribute 'diagnosis', which is utilized as an output class of the prediction. It contains the class either benign or malignant. The

exact distribution of patients belonging to the 'benign' or 'malignant' is shown in Fig. 1. Fig. 2 shows the overall understanding of the dataset.

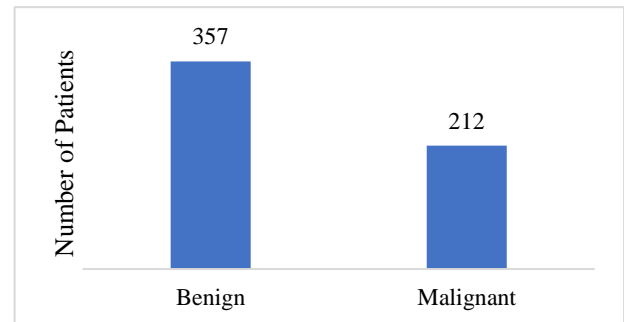


Fig. 1. Distribution of benign and malignant patients.

V. THE 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 the 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 steps are executed in order to obtain the task of prediction, estimation, etc. The following section explains application of data mining techniques for breast cancer prediction. A detailed explanation of data pre-processing techniques is provided. Then, the combinations of different RNN based architecture are described and compared for identifying significantly better prediction performance.

A. Dataset Pre-Processing

Once data are collected, pre-processing techniques are applied in order to obtain a balanced dataset. Pre-processing techniques include checking and handling missing values, scaling some attributes. Existing 'nan' values are checked for each attribute. These values are replaced with mean values for the corresponding attribute. Some of the attributes such as 'id', 'Unnamed :32' are eliminated since they have no contribution in prediction. Next, feature scaling of relevant attributes is performed, which enhances the efficiency while fitting to a classifier. The dataset contains 10 real-valued attributes such as, radius (mean of distances from centre to points on the perimeter), texture (standard deviation of grey-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). For implementing feature scaling operation, these attribute values are scaled down into a specified range from 0 to 1. Applying these pre-processing techniques will yield a transformed dataset that can be fitted to the classifier. The transformed dataset is partitioned into a training set and testing dataset, which is obtained by partitioning the transformed dataset with the ratio of 7:3. Training data are fitted to RNN models and predictions are received for the test dataset.

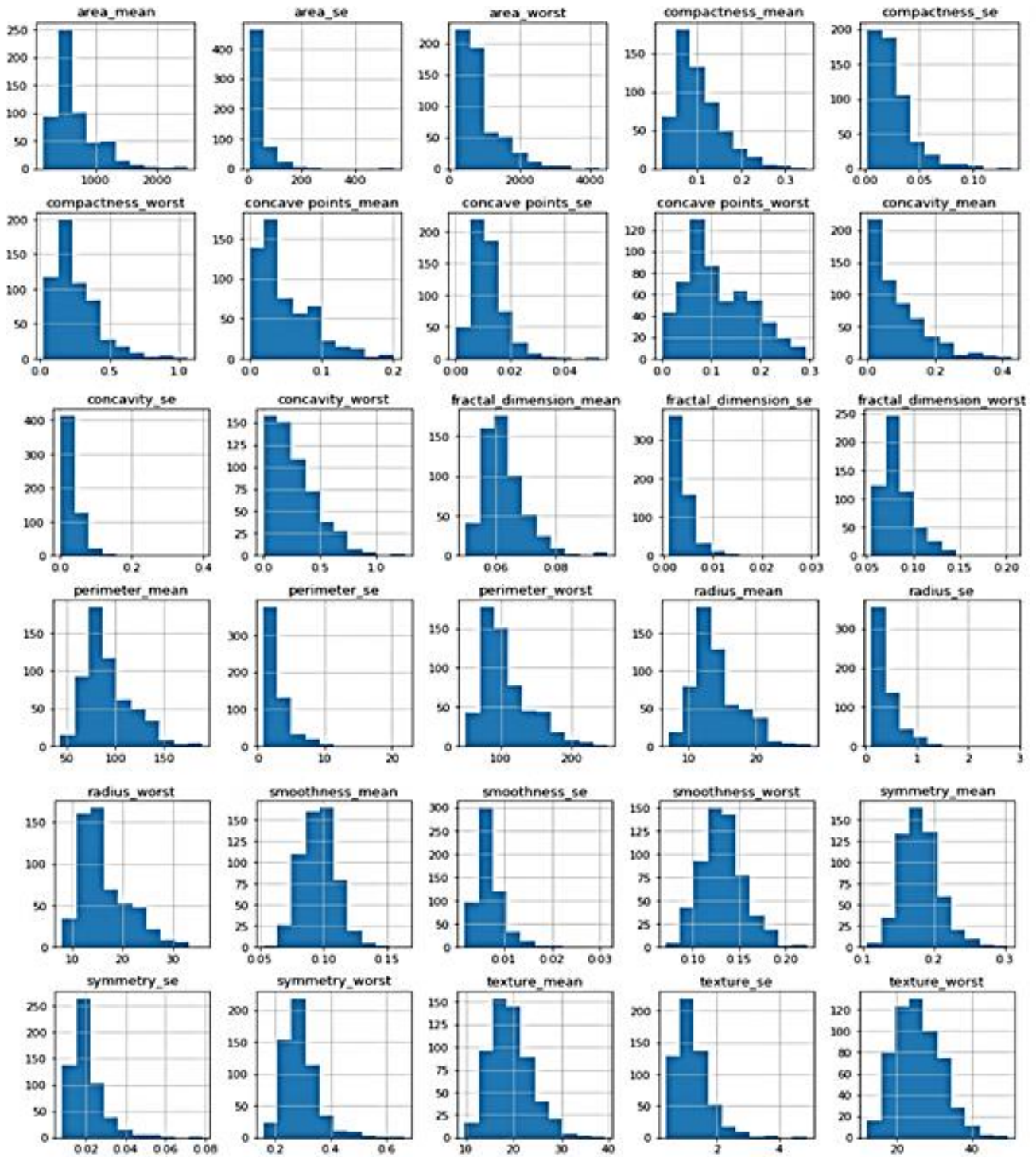


Fig. 2. Visual representation of dataset attributes.

B. Methodology and Implementation

The main objective of the proposed classifier is to predict whether a patient has breast cancer disease or not using deep learning techniques. Deep learning techniques assist in recognising features automatically from raw data using a supervised learning paradigm by an end-to-end training

procedure. The classification process, supervised learning technique, aims to detect benign or malignant cancer patients. The proposed method implements GRU and LSTM-BRNN based framework for such predictions.

The present paper proposes a stacked GRU-LSTM based model that contains an 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 maximised efficiency. This section describes the specification of the model along with its hyper-parameters. The proposed model consists of four LSTM and GRU layers with 128, 64, 32, 16 units, respectively. Each of these layers is followed by a layer having a dropout rate of 20 %. Next, four dense layers are stacked in this model with 8, 4, 2, 1 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 the best predictive results. The neural network receives a total of 305 785 parameters and trains those parameters in order to obtain prediction. The detailed description of the model is given in Table I.

TABLE I
DESCRIPTION OF STACKED GRU-LSTM-BRNN MODEL

Layers	Number of nodes / dropout rate	Output shape	Number of received parameters
Bidirectional GRU	128	(None, 30, 256)	99 840
Dropout	0.2	(None, 30, 256)	0
Bidirectional GRU	64	(None, 30, 128)	164 352
Dropout	0.2	(None, 30, 128)	0
Bidirectional LSTM	32	(None, 30, 64)	30 912
Dropout	0.2	(None, 30, 64)	0
Bidirectional LSTM	16	(None, 32)	10 386
Dropout	0.2	(None, 32)	0
Dense	8	(None, 8)	264
Dense	4	(None, 4)	36
Dense	2	(None, 2)	10
Dense	1	(None, 1)	3

C. Baseline Classifiers and Implementation

The proposed model is evaluated against a set of other RNN models such as simple RNN, stacked LSTM model, and stacked GRU model. The specified models are elaborated with their corresponding description and implementation details.

1. 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 layer with a rate of 20 %. However, these layers are followed by incorporating four dense layers. First four RNN layers and last dense layer receive 'sigmoid' as an activation function. A total of 33 065 parameters are fed into this model during the training phase. Summarised description of this model is given in Table II.

TABLE II
DESCRIPTION OF STACKED SIMPLE RNN MODEL

Layers	Number of nodes / dropout rate	Output shape	Number of received parameters
Simple RNN	128	(None, 30, 128)	16 640
Dropout	0.2	(None, 30, 128)	0
Simple RNN	64	(None, 30, 64)	12 352
Dropout	0.2	(None, 30, 64)	0
Simple RNN	32	(None, 30, 32)	3104
Dropout	0.2	(None, 30, 32)	0
Simple RNN	16	(None, 16)	784
Dropout	0.2	(None, 16)	0
Dense	8	(None, 8)	136
Dense	4	(None, 4)	36
Dense	2	(None, 2)	10
Dense	1	(None, 1)	3

2. Stacked LSTM-RNN Model

A stacked LSTM model is implemented that contains four LSTM layers, each of them is followed by dropout layers. Again, four dense layers are also incorporated into the model. LSTM layers and the last dense layer are implemented using the '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 results. This model is described in Table III.

TABLE III
DESCRIPTION OF STACKED LSTM-RNN MODEL

Layers	Number of nodes / dropout rate	Output shape	Number of received parameters
LSTM	128	(None, 30, 128)	66 560
Dropout	0.2	(None, 30, 128)	0
LSTM	64	(None, 30, 64)	49 408
Dropout	0.2	(None, 30, 64)	0
LSTM	32	(None, 30, 32)	12 416
Dropout	0.2	(None, 30, 32)	0
LSTM	16	(None, 16)	3136
Dropout	0.2	(None, 16)	0
Dense	8	(None, 8)	136
Dense	4	(None, 4)	36
Dense	2	(None, 2)	10
Dense	1	(None, 1)	3

3. Stacked GRU-RNN Model

Like the other two baseline classifiers, this model also consists of sequences of GRU as well as 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 an activation function just like the previous two models. The model is capable of drawing prediction by considering 98 825

parameters during the training phase. Brief description of the model is shown in Table IV.

TABLE IV
DESCRIPTION OF STACKED GRU-RNN MODEL

Layers	Number of nodes / dropout rate	Output shape	Number of received parameters
GRU	128	(None, 30, 128)	49 920
Dropout	0.2	(None, 30, 128)	0
GRU	64	(None, 30, 64)	37 056
Dropout	0.2	(None, 30, 64)	0
GRU	32	(None, 30, 32)	9312
Dropout	0.2	(None, 30, 32)	0
GRU	16	(None, 16)	2352
Dropout	0.2	(None, 16)	136
Dense	8	(None, 8)	36
Dense	4	(None, 4)	10
Dense	2	(None, 2)	3
Dense	1	(None, 1)	136

D. General Model Structure

This section provides the general structure of RNN model architecture. All the aforementioned models are implemented with the same number of layers, along with the 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 is different for each model. The implemented models follow the same structure in terms of the number of nodes present in each layer, present dropout layer rate, epochs, batch size, and optimizer. This will provide a similar platform to compare models in terms of the prediction. The specified parameters common to each of the implemented models are defined in Table V.

TABLE V
GENERAL STRUCTURE IMPLEMENTED FOR RNN-BASED MODEL

Layer number >	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							

VI. EXPERIMENTAL RESULTS

During the training process, RNN models are learned by feeding a training dataset. Training process is evaluated in terms of loss and accuracy over each epoch. The proposed model along with other baseline classifiers are trained through certain epoch sizes. Initially epoch size was chosen as 20, 50, and 100. All the models were trained using these specified epoch sizes.

After training, models were evaluated and compared with respect to testing accuracy. As shown in Table VI, epoch size of 100 shows the best prediction efficiency. Hence, epoch size of 100 is considered to be the best training criterion. Accuracy and loss acquired through each of the 100 epochs for each of the four models are indicated in Figs. 3–6. Using keras [26] with TensorFlow [27] backend as deep learning framework, all the models are implemented. The experiments were carried out on Windows 10 machine with Intel Core i5-9300H CPU, NVIDIA GTX 1650 GPU (4 GB memory).

TABLE VI
PREDICTION ACCURACY OF ALL RNN-BASED MODELS FOR DIFFERENT EPOCH SIZE

Epoch size	Simple RNN model	Stacked LSTM model	Stacked GRU model	Proposed stacked GRU-LSTM model
20	68.62 %	68.69 %	70.82 %	84.57 %
50	71.81 %	80.85 %	84.04 %	93.06 %
100	77.13 %	84.57 %	93.62 %	97.34 %

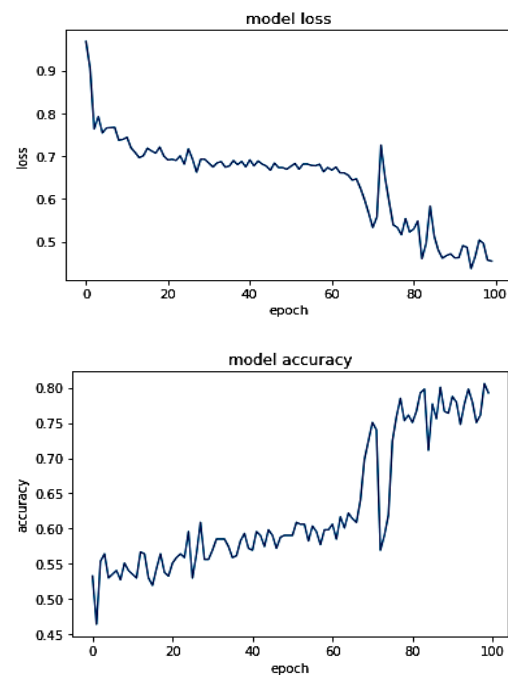


Fig. 3. Loss and accuracy plot of stacked simple-RNN model obtained over epochs.

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 models, stacked LSTM model, and stacked GRU model. On completion of the training procedure, test accuracy is measured after 100th epoch. In other words, 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

instantiate the evaluation of deep models with respect to employed metrics. The comparative study is shown in Table VII. From the comparative study it is clear that the proposed model indicates much better promising result over other classifiers.

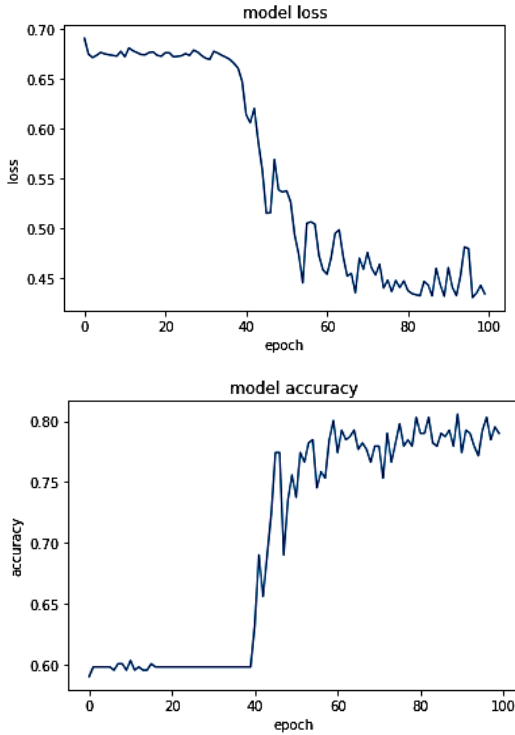


Fig. 4. Loss and accuracy plot of stacked LSTM model obtained over epochs.

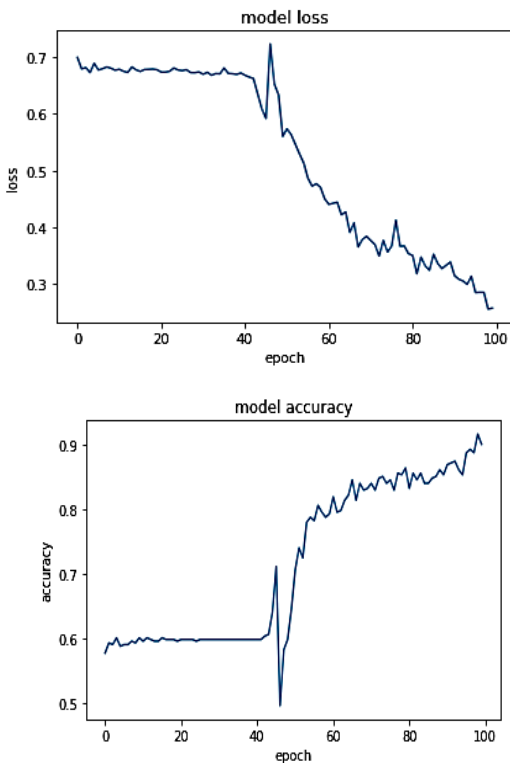


Fig. 5. Loss and accuracy plot of stacked GRU model obtained over epochs.

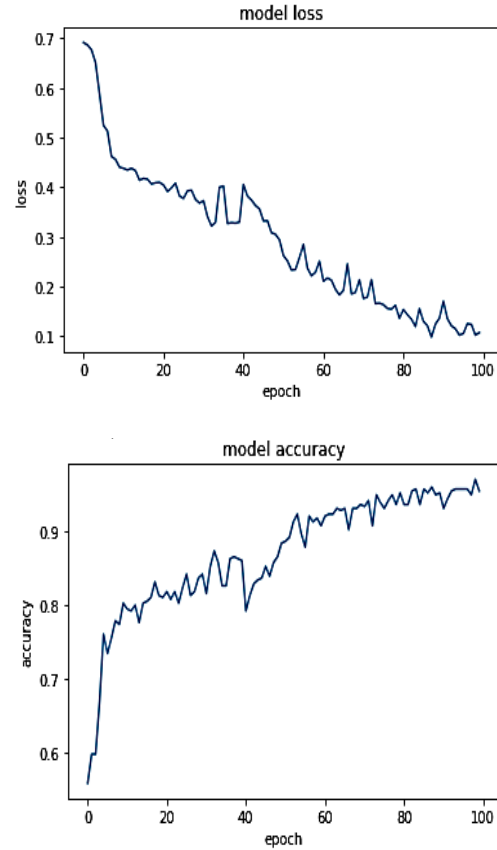


Fig. 6. Loss and accuracy plot of the proposed stacked GRU-LSTM model obtained over epochs.

TABLE VII

PERFORMANCE SUMMARIZATION OF ALL RNN-BASED MODELS

Evaluating metrics	Simple RNN model	Stacked LSTM model	Stacked GRU model	Proposed stacked GRU-LSTM model
Test loss	0.4586013 99305019	0.374331624 5839951	0.21349462 422918766	0.09276566210 579365
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

As shown in Table VII, stacked LSTM deep model provides a better classification result than the 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, the improvement of the classification result is observed in the stacked GRU deep model. By incorporating the advantages of LSTM as well as GRU into a single platform, even more improvised results are obtained in terms of classification. Hence, this model is favoured as the best predictive model for BC detection with the highest efficiency.

VII. CONCLUSION

Breast cancer is a severe disease that needs to be handled carefully. Detection of this disease at an early stage is quite helpful in saving patients' lives. The objective of this study is to detect the feasibility of utilising previous medical records and to determine the probability of being affected by breast cancer disease. Using deep learning methods, a stacked GRU-LSTM layer based model has been proposed and implemented in this paper. Interfering attributes that have impact on this disease have been considered while designing the model with necessary parameter tuning. The highest value of accuracy, F1-score, Cohen-kappa score and the lowest value of test loss and MSE have been attained by the stacked GRU-LSTM model which denotes superiority of the model over other baseline classifiers. The proposed method achieves promising results 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," *Surgery (Oxford)*, vol. 34, no. 1, pp. 25–31, Jan. 2016. <https://doi.org/10.1016/j.mpsur.2015.10.005>
- [2] J. Sathwara, S. Bobdey, and B. Ganesh, "Breast cancer survival studies in India: A review," *International Journal of Research in Medical Sciences*, vol. 4, no. 8, pp. 3102–3108, Aug. 2016. <https://doi.org/10.18203/2320-6012.ijrms20162266>
- [3] C. K. Anders, R. Johnson, J. Litton, M. Phillips, and A. Bleyer, "Breast cancer before age 40 years," *Seminars in Oncology*, vol. 36, no. 3, pp. 237–249, Jun. 2009. <https://doi.org/10.1053/j.seminoncol.2009.03.001>
- [4] Ó. Marbán, G. Mariscal, and J. Segovia, "A data mining & knowledge discovery process model," in *Data Mining and Knowledge Discovery in Real Life Applications*, (J. Ponce and A. Karahoca, Eds.). In Tech, 2009. <https://doi.org/10.5772/6438>
- [5] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, vol. 19, pp. 221–248, Jun. 2017. <https://doi.org/10.1146/annurev-bioeng-071516-044442>
- [6] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Physica D: Nonlinear Phenomena*, vol. 404, article number 132306, Mar. 2020. <https://doi.org/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," 2014. [Online]. Available: <https://arxiv.org/abs/1412.3555>
- [8] D. Soutner and L. Müller, "Application of LSTM neural networks in language modelling," in Habernal I., Matoušek V. (eds) Text, Speech, and Dialogue. TSD 2013. Lecture Notes in Computer Science, vol. 8082. Springer, Berlin, Heidelberg. https://doi.org/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," *Computers & Electrical Engineering*, vol. 70, pp. 871–882, Aug. 2018. <https://doi.org/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," in *2015 IEEE International Conference on Computational Intelligence and Computing Research*, IEEE, 2016. <https://doi.org/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 Systems with Applications*, vol. 120, pp. 103–115, Apr. 2019. <https://doi.org/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 Systems with Applications*, vol. 117, pp. 103–111, Mar. 2019. <https://doi.org/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, Dec. 2018. <https://doi.org/10.1016/j.ict.2018.10.007>
- [14] S. Kaymak, A. Helwan, and D. Uzun, "Breast cancer image classification using artificial neural networks," *Procedia Computer Science*, vol. 120, pp. 126–131, 2017. <https://doi.org/10.1016/j.procs.2017.11.219>
- [15] V. Chaurasia and S. Pal, "Data mining techniques: To predict and resolve breast cancer survivability," *International Journal of Computer Science and Mobile Computing*, vol. 3, no. 1, pp. 10–22, Jan. 2014.
- [16] S. A. Medjahed, T. A. Saadi, and A. Benyettou, "Breast cancer diagnosis by using k-nearest neighbor with different distances and classification rules," *International Journal of Computer Applications*, vol. 62, no. 1, pp. 1–5, Jan. 2013. <https://doi.org/10.5120/10041-4635>
- [17] D. Delen, G. Walker, and A. Kadam, "Predicting breast cancer survivability: A comparison of three data mining methods," *Artificial Intelligence in Medicine*, vol. 34, no. 2, pp. 113–127, Jun. 2005. <https://doi.org/10.1016/j.artmed.2004.07.002>
- [18] L. G. Ahmad, A. T. Eshlaghy, A. Poorebrahimi, M. Ebrahimi, and A. R. Razavi, "Using three machine learning techniques for predicting breast cancer recurrence," *Journal of Health & Medical Informatics*, vol. 4, no. 2, 2013. <https://doi.org/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," 2018. [Online]. Available: <https://arxiv.org/abs/1811.03378>
- [20] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A critical review of recurrent neural networks for sequence learning," 2015. [Online]. Available: <https://arxiv.org/abs/1506.00019>
- [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," *Applied Soft Computing*, vol. 85, article number 105765, Dec. 2019. <https://doi.org/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," in *International Conference for High Performance Computing, Networking, Storage and Analysis*, 2019. <https://doi.org/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. <https://doi.org/10.4467/20838476SI.16.004.6185>
- [24] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations*, 2015. [Online]. Available: <https://arxiv.org/abs/1412.6980>
- [25] D. Dua, and C. Graff, *UCI Machine Learning Repository*. Irvine, CA: University of California, School of Information and Computer Science. [Online]. Available: <https://archive.ics.uci.edu/ml>
- [26] F. Chollet et al., *Keras*, 2015. [Online.] Available: <https://keras.io>
- [27] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: A system for large-scale machine learning," in *12th USENIX Symposium on Operating Systems Design and Implementation*, 2016, pp. 265–283.



Shawni Dutta has been working as a Full-Time Teacher in the Computer Science department of Bhawanipur Education Society College, Kolkata since March 2019. She has completed her bachelor's and master's degree from the University of Calcutta in the subject of Computer Science. Her research areas are Machine Learning, Medical Data Analysis, Malaria, and Recommended systems. She has

published a book on COVID-19- Validation of Disease by Machine Learning Approach and the book is published by LAP, Lambert Academic Publishing, Eastern Europe. Prof. Dutta presented Invited Speech on "Growth of COVID-19 in the World using Machine Learning Approach," in Virtual Summit on Saturday, July 11th, 2020, organized by Sciinov Group, Europe. Prof. Dutta has also published Book Chapters on Artificial Intelligence Based Diagnosis of Brain Tumor Diseases by IOP Publishers. She has written book chapter Heterogeneity Aspects of Breast for Breast Cancer Patients Change due to COVID-19, Computational Intelligence for Managing Pandemics, CIMP (COVID-19) – 2020. All total, she has published 25 research papers in International Journals.

E-mail: shawnidutta83@gmail.com

ORCID iD: <https://orcid.org/0000-0001-8557-0376>



Jyotsna Kumar Mandal has completed his M. tech. in computer science from the University of Calcutta in 1987. He has been awarded PhD (Engineering) in computer science & engineering by Jadavpur University in 2000. Presently he is working as a professor of computer science & engineering. He was former Dean, Faculty of Engineering, Technology & Management, Kalyani

University for two consecutive times (2008–2012). He is Director of IQAC, Kalyani University and Chairman of CIRM and Placement Cell. Prof. Mandal has 30 years of teaching and research experience in coding theory, data and network security and authentication; remote sensing & GIS-based applications, data compression, error correction, image processing, visual cryptography, and steganography. Twenty-Three Research Scholars awarded PhD Degrees under his guidance.

E-mail: jkm.cse@gmail.com

ORCID iD: <https://orcid.org/0000-0001-9447-647X>



Tai-Hoon Kim has received the B.E. and M.E. degrees from Sungkyunkwan University, South Korea, and the PhD degrees from the University of Bristol, UK, and from the University of Tasmania, Australia. He is currently with the Department of Convergence Security, Sungshin Women's University, South Korea. He has published 300 research papers in international journals and conferences. His main

research interests include security engineering for IT products, IT systems, development processes, and operational environments. He is an Editor of the Elsevier, Springer, and many more journals.

E-mail: taihoonn@sungshin.ac.kr

ORCID iD: <https://orcid.org/0000-0001-7823-5457>



Samir Kumar Bandyopadhyay is a Professor in the Department of Computer Science and Engineering in the Bhawanipur Education Society College, India. He did his Bachelor's and Master's Degree in Engineering Degree and Medical Science from the University of Calcutta. Prof. Bandyopadhyay did three PhD in Medical Science. First, he did his PhD in Cardiology at the University of Calcutta. Prof.

Bandyopadhyay also did his second PhD on Breast Cancer from Bristol University, UK. Finally, he obtained his PhD degree in Brain Tumor from MIT, USA. He served as Registrar of the University of Calcutta and Vice-Chancellor of the West Bengal University of Technology in West Bengal, India.

E-mail: 1954samir@gmail.com

ORCID iD: <https://orcid.org/0000-0002-4868-3459>