Journal of Information Systems Engineering and Management
2025, 10(32s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

ConvGateRNN: Convolutional Gated Attention Recurrent Neural Network for Cervical Cancer Screening

Anusha R[1*], Dr Srinivas Prasad[2]
[1] Research Scholar, Department of Computer Science and Engineering, GITAM Deemed to be University, Andhra Pradesh, India. amanuknr@gmail.com
[2] Professor, Department of Computer Science and Engineering, GITAM Deemed to be University, Andhra Pradesh, India. sprasad@gitam.edu

ARTICLE INFO

Received: 24 Dec 2024
Revised: 12 Feb 2025
Accepted: 26 Feb 2025

ABSTRACT

Cervical cancer continues to pose a substantial obstacle to global health, especially in areas where access to healthcare is restricted. Traditional screening methods are often manual, laborious, and prone to error. In this paper, a novel method for early-stage cervical cancer detection with a Convolutional Gated Attention Recurrent Neural Network (ConvGateRNN) is proposed. The methodology involves preprocessing and feature selection using the spotted hyena optimization technique, followed by the implementation of a ConvGateRNN classifier. The ConvGateRNN leverages deep learning techniques and attention mechanisms to effectively identify disease traits from cervical cancer datasets. Experimental findings showcase ConvGateRNN' s exceptional efficacy relative to current approaches, achieving a remarkable accuracy of 96.18% alongside noteworthy precision, recall, and F1-score metrics. The promising results suggest that ConvGateRNN offers a favorable solution for unerring and efficient cervical cancer screening, potentially revolutionizing current practices and improving healthcare outcomes for women worldwide.

Keywords: Cervical Cancer, Preprocessing, Hyena Optimization, Gated Recurrent Neural Network, Classification

INTRODUCTION

The cervix, situated beneath the fibromuscular region of the uterus, is a crucial component of the biological system in females responsible for reproduction. [1]. Composed of elongated columnar cells and thin, scale-like squamous cells, it houses the transformation zone where normal cells transition into abnormal ones [2]. This zone's location varies among individuals, leading to cervix classification into types 1, 2, and 3 based on this variation. Aberrant cells originating in the transformation zone possess the potential to evolve into cervical cancer over time [3]. Globally, cervical cancer is positioned as a significant factor that contributes to the development of cancer -related deaths in women, with a substantial majority of cases (80%) occurring in developing countries [4]. According to data from the Human Papillomavirus (HPV) Information Centre, Ethiopia reports a yearly diagnosis of new cervical cancer cases. The incidence and mortality rates of cervical cancer are expected to be higher in remote areas characterized by poor healthcare infrastructure, limited access to medical services, and a scarcity of trained medical professionals [5]. Various screening approaches for cervical cancer include visual inspection subsequent to the administration of acetic acid (VIA), human papillomavirus (HPV) test and Papanicolaou test (Pap). The types of cervical cancer are identified by microscopic visual inspection of histopathological characteristics. These manual procedures are often laborious, time-consuming, and susceptible to mistakes owing to variations between and between observers [6]. The use of computer-aided diagnostic (CAD) methods has promise in the automation of manual screening and diagnosis operations, hence assisting medical practitioners in various decision-making processes such as illness identification, cervix localization, cancer grading, and treatment planning [7]. Deep learning has improved computer vision, language processing, and predictive analytics[8]. The multi-layer perceptron (MLP) neural network system of deep learning has the capability to acquire more abstract characteristics in attributes, which is anticipated to effectively tackle the challenges faced by traditional medical CAD systems [9]. Nevertheless, it is vital to reinforce the deep learning procedures using a comprehensive database, particularly for instances that provide great results. Numerous transfer learning and ensemble learning methodologies have been examined in prior research [10] as potential solutions to address this challenge. The CNN is used inside a CAD framework to effectively detect cervical cancer in the context of urban healthcare in smart cities. Given these circumstances, the present study makes the following contributions:

- Feature selection is a fundamental procedure that involves the identification and removal of essential elements from a provided dataset. Its purpose is to decrease computational demands, reduce dimensionality, and enhance prediction accuracy. In this, we implemented the spotted hyena optimization technique. The system employs the hunting mechanism to continuously update the location of hyenas in a search space.
- The exploration of an RNN-based technique is under underway in order to acquire representations of features related to cervical cancer. The neuron activations enable the exact detection of contaminated areas.

Here is the outline of the paper: There is an introduction in Section 1. The literature review is summarized in Section 2. The system's operating procedures and execution are explained in detail in Section 3, while Section 4 presents the results and accomplishments that were attained. In Section 5, we draw a conclusion and discuss some potential future study avenues

RELATED WORKS:

The goal of implementing automated cervical cytology screening is to enhance the overall efficacy of cervical cancer screening initiatives, weaken the occurrence and fatality rates related to cervical cancer, and ultimately enhance the health outcomes of women. Moreover, automated solutions have the potential to decrease healthcare expenses linked to cervical cancer screening and treatment, hence enhancing accessibility for women in settings with low resources. Autonomous diagnosis and reporting of findings will be made possible by future automated screening systems via the use of advanced methods such as medical imaging, computer vision, and machine learning. The following is a compilation of extant literature focusing on identifying the cervical cancer.

Initially, the data was understood by the use of visualization methods in reference [11]. Afterwards, survival prediction models for classification and regression were created using ML algorithms. The duration between the initial diagnosis and the patient's passing was trained into the algorithms of the classification models. We were able to learn a lot about the model's functioning and reasoning process by using attribute weights, which highlighted elements that had a big influence on predictions. The accuracy rate attained by the gradient boosting trees technique was 81.55%.

Presented in [12] is an innovative method that exploits a stacked ensemble voting classifier—a combination of three machine learning models—and a KNN Imputer to manage the missing data. A remarkable set of metrics, including an F1 score of 0.97, recall of 0.96, precision of 0.98, and accuracy of 0.9941, are produced by the suggested model. There were a number of different classification algorithms employed in [13], including logistic regression, MLP, SVM, KNN, and a few naive Bayes approaches. It is possible to create a superfluous system that can employ alternative ways to corroborate the result supplied by the original algorithm, given the high classification performances attained with numerous AI algorithms and multiple class-balancing strategies. We employed Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), Multi-layer Perceptron (MLP), and Naive Bayes (NB) as machine learning classification practices to assess the imbalanced input and target datasets in [14]. Between 88% and 94% of the time, it was accurate (84% to 89% of the time). As mentioned in [15], Several popular ML classifiers, including Naïve-Bayes, Logistic Regression, K-Nearest-Neighbor, Support Vector Machine (SVM), Linear Discriminant Analysis, Multi-Layer Perceptron, Decision Tree (DT), and Random Forest (RF), have been tested on both untouched and standardized data, using MinMax, Standard, and Normalization as scaling methods. The results show promise for these programs. By using feature selection approaches namely Recursive feature elimination (RFE), Univariate feature-selection and optimization opportunities are investigated for the supreme three ML algorithms—RF, SVM, and DT—obtained in cervical cancer detection. Among all the current implementations, Random Forest predictor alongside RFE (RF-RFE) outperforms them all. Three feature modification techniques were used on these datasets in [16]: logarithmic function, Z-score and sine function,. We tested the categorization abilities of many supervised machine learning systems. While a Random Forest (RF) algorithm produced the greatest results for Hinselmann (99.16%) and Schiller (98.58%), respectively, a Random Tree (RT) strategy produced the best results for cytology (98.65%) and biopsy (98.33%) data. Logarithmic feature transformation surpassed others on biopsy datasets, whereas sine function displayed better on cytology datasets. An effective model for prediction and feature selection was suggested for cervical cancer datasets in [17] that tackles this difficulty by using Boruta analysis and the SVM approach. The technique of Boruta analysis is used. To find feature subsets from the source data that are important to assigned categorization activity, it primarily uses a better version of the random forest approach. Finding out which cervical cancer screening parameters are most important for identifying high-risk individuals is the main goal of the suggested methodology. Employing a variety of machine learning feature selection techniques, the authors of [18] determined the leading causes of cervical cancer. Using all characteristics in addition to these crucial features, we trained several machine learning algorithms, namely Ensemble Learners, Decision Trees (DTs), and Support Vector Machines (SVMs), after this selection. With 96% accuracy, AdaBoost can divide this imbalanced dataset into healthy and harmful categories.

Several machine learning algorithms have been employed to address regression, multi- or single-class classification, and outlier identification in problem statements. Two main steps comprise support vector machines and linear regression: determining the hyperplane's location and solving a quadratic equation. During the beginning, this procedure made use of well-established, pre-existing quadratic optimization techniques. Small data sets with around a thousand samples were ideal for this method. Their great numerical accuracy and lack of requirement-specific design make them advantageous, since they need minimal developer involvement. When confronted with larger datasets and memory constraints that render the kernel unfit, a more customized approach must be adopted. This approach should encompass a broader spectrum of factors, catering to the specific needs and limitations of the dataset and computational resources available. Deep learning has the power to revolutionize cancer prediction by making use of massive amounts of data. Furthermore, most existing techniques depend on deep learning models that consist of several deep layers and a large number of parameters, requiring significant computer power and lengthy training periods. Moreover, most optimization techniques rely on extracting deep features from a solitary CNN. Nevertheless, the retrieval of deep features from CNNs with diverse architectures is seen more advantageous. The Spotted Hyena Feature Selection Method based on CNN-GRU presents an elegant solution to several challenges associated with deep learning models, particularly amidst the context of large-scale data like cancer prediction. By reducing the computational burden, improving memory efficiency, and enhancing generalization, this hybrid approach enables deep learning models to operate more efficiently, making it suitable for real-world applications where resources are constrained.

PROPOSED METHODOLOGY

This study focuses on recognizing cervical cancer at initial stage using classification algorithms applied to several prominent aspects within the dataset. The proposed methodology is comprised of the subtasks illustrated as in Figure 1.

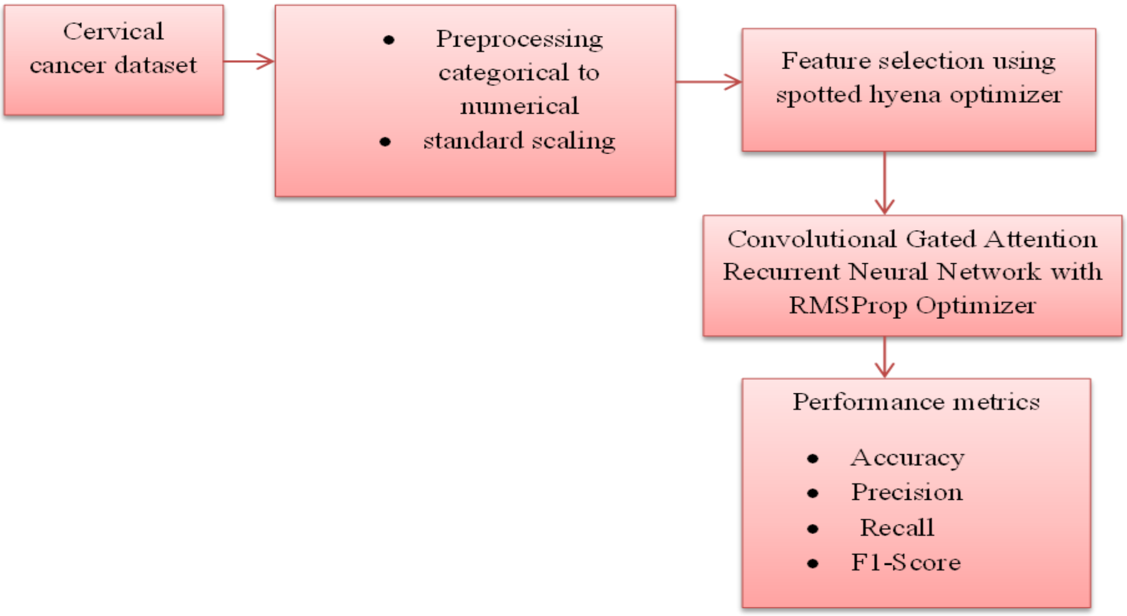


Figure 1: Block diagram for cervical cancer detection

Certain critical phases of this methodology require additional attention: (1) pre-processing and feature selection; and (2) model construction and analysis. In detail, each stage is described below.

3.1 Dataset description:

The dataset [19] was obtained from 'Hospital Universitario de Caracas' located in Caracas, Venezuela. The data collection includes 858 patient's demographics, lifestyle choices, and medical history. Concerns about their privacy led a number of patients to opt out of answering certain questions (missing values).Table 1 is a glimpse of the cervical cancer dataset.

Table 1. Cervical cancer dataset

Variable name	Role	Type	Missing value
age	feature	integer	no
Number of sexual partners	feature	continuous	yes
First sexual intercourse	feature	continuous	yes
Num of pregnancies	feature	continuous	yes
Smokes	feature	continuous	yes
Smokes (years)	feature	continuous	yes
Smokes(packs/ years)	feature	continuous	yes
Hormonal contraceptives	feature	continuous	yes
Hormonal contraceptives (years)	feature	continuous	yes
IUD	feature	continuous	yes

3.2 Preprocessing of data

Data preprocessing refers to the procedure of preparing and refining raw data, converting the information given in the dataset to streamline the acquisition and processing of information. Typically, the dataset exhibits a substantial disparity between its maximum and minimum values; therefore, the process of data normalization serves to reduce the intricacy of the algorithm associated with its processing. Each value of the categorical variable is mapped to a distinct feature variable by a single hot encoder. If the categorical variable of that data object contains the value of the new feature variable and all other feature variables receive the value zero, then the value of the new feature variable is set to 1. The first heated encoding is a simple encoding method that is well-suited for categorical encoding in situations where the categorical variable contains a relatively small number of unique values. A heated encoder can generate a significant number of new variables equal to the count of unique values for the categorical variable, which can lead to critical memory issues. The utilization of a single hot encoder to convert the categorical values of the variable 'Grade'. By means of standard scaling, the training data is smoothed and made ready for modelling. Feature scaling or data normalization is a commonly used technique in the field of data processing to standardize the range of independent variables or data attribute values. Therefore, to guarantee that each characteristic has an equitable impact on the ultimate outcome before using classification algorithms for early-stage cervical cancer diagnosis, we use Standard Scaling to scale our data. The Standard Scaler formula is calculated as $\frac{X_i - X_{mean}}{X_{std}}$, which means it adjusts the mean to a value of 0.

$$X_{new} = \frac{X_i - X_{mean}}{standard\ deviation} \quad (1)$$

The StandardScaler is susceptible to the influence of outliers due to their impact on the mean. Following the deployment of StandardScaler, our data exhibits a clustering around 0, however it remains right-skewed and has a significant scale, ranging from -1 to 30. The data exhibit clustering and lack significant outliers. StandardScaler yields unsatisfactory outcomes when higher values are present as outliers. Outliers have a significant impact on all scaling data. Furthermore, it had an impact on this data. In instances of this kind, it is advisable to first eliminate outliers.

3.3 Selection of feature using spotted hyena

The process of optimization involves the identification of the optimal combination of hyperparameters. Additionally, the approach involves identifying the optimal model structure and training constraints to maximize the performance of the archetype across both validation and testing datasets. This ensures that the model achieves robust and reliable results, generalizing well to unseen data. The process of feature selection involves the identification and selection of a subgroup of features from the dataset, with the aim of effectively eliminating various noises, missing values, and redundant features. Several feature selection techniques are available, filtering, wrapping, and embedding approaches. Filter techniques choose the features explicitly, without the utilization of any learning process. This approach assigns a rating to each feature based on either univariate or multivariate metrics, and then chooses the characteristics with the highest ranking. The present study employs a selection strategy based on spotted hyenas to identify a subset of attributes from the dataset pertaining to cervical cancer.

Spotted Hyenas live in a collective of over 100 individuals. Typically, they engage in collective hunting behavior. The classification of hyenas encompasses four distinct varieties, namely spotted, striped, brown, and aardwolf. Spotted hyenas are a large species of hyenas that are very skilled predators. They emit a sound like human laughter as a means of communication. Their body is covered with patches. Hyenas establish tight groups to facilitate cooperation among themselves. Below is a description of the mathematical modelling of SHO. SHO consists of three essential stages: surrounding, hunting, and attacking prey. The present optimum solution is something that closely approximates the best option. The remaining spotted hyenas attempt to modify their postures subsequent to the identification of the optimal option. The mathematical formulation of the encircling behavior shown by spotted hyenas is expressed via the use of the following equations:

$$\overrightarrow{D_h} = \left|\vec{Y}.\overrightarrow{Q_q}(s) - \vec{Q}(s)\right| \tag{2}$$

$$\vec{Q}(s+1) = \overrightarrow{Q_q}(s) - \vec{Z}.\overrightarrow{D_h} \tag{3}$$

Where, $\overrightarrow{D_h}$ represents the distance between the desired prey and the hyena. The coefficient vectors are denoted as $\vec{Y}$and $\vec{Z}$. The coefficient-vectors represent the instantaneous values. The location-vector of the prey is denoted as $\overrightarrow{Q_q}$, whereas the position vector of the spotted hyena is represented as $\vec{Q}$. The computations for $\vec{Y}$ and $\vec{Z}$ are as follows:

$$\vec{Y} = 2.\overrightarrow{rand_1} \tag{4}$$

$$\vec{Z} = 2\vec{h}.\overrightarrow{rand_1} - \vec{h} \tag{5}$$

$$\vec{h} = 5 - \left(iteration \times \left(\frac{5}{max_{iteration}}\right)\right) \tag{6}$$

Where, $iteration = 1,2,3, \dots max_{iteration}$ (7)

Here, $\vec{h}$ exhibits a linear drop from 5 to 0 as the iterations go. The purpose of this mechanism is to ensure the equilibrium between exploration and exploitation, where $\overrightarrow{rand_1}$and $\overrightarrow{rand_2}$ represent the random vectors inside the interval [0,1]. The values of $\vec{Y}$and $\vec{Z}$ are modified in order to facilitate the movement of spotted hyenas to various regions in relation to their present location. Spotted hyenas possess the ability to engage in random positional updates around their prey via the use of (2) and (3). To replicate the hunting behaviour of spotted hyenas, we use the assumption that the optimal search agent has information on the whereabouts of its prey. The remaining search agents establish a group of reliable acquaintances and adjust their position based on the most optimal search agent. The hunting mechanism may be mathematically expressed using the following equations:

$$\overrightarrow{D_h} = \left|\vec{Y}.\overrightarrow{Q_h} - \overrightarrow{Q_k}\right| \tag{8}$$

$$\overrightarrow{Q_k} = \overrightarrow{Q_h} - \vec{Z}.\overrightarrow{D_h} \tag{9}$$

$$\overrightarrow{C_h} = \overrightarrow{Q_k} + \overrightarrow{Q_{k+1}} + \dots + \overrightarrow{Q_{k+N}} \tag{10}$$

Where, $Q_h$represents the location of the first best spotted hyena, whereas $Q_k$represents the location of other spotted hyenas. The count of spotted hyenas, signified as N, may be identified using the following calculation:

$$N = count_{NS}(\overrightarrow{Q_h}, \overrightarrow{Q_{h+1}}, \overrightarrow{Q_{h+2}}, \dots (\overrightarrow{Q_h} + \vec{M})) \tag{11}$$

Where random vector $\vec{M}$ is defined within the range of [0.5, 1]. The variable NS represents the quantity of candidate solutions that exhibit similarity to the optimum solution inside the specified search space. The cluster of N optimum solutions is denoted as $\overrightarrow{C_h}$. The mathematical expression for attacking the prey is shown below.

$$\vec{Q}(s+1) = \frac{\overrightarrow{C_h}}{N} \tag{12}$$

Where, $\vec{Q}(s+1)$ is used to hold the optimal solution, while the remaining search agents adjust their locations based on the position of the optimal search agent. The SHO provide their hyenas the ability to modify their locations and launch attacks on their victims. Nevertheless, the dataset contains a significant number of duplicated and irrelevant characteristics, which provide challenges for the classification method in developing efficient anomaly detection classifiers. The use of the spotted hyena optimization technique is employed to enhance the efficacy of classifiers through the exploration of the most optimum subset of characteristics. The process involves the independent selection of redundant features, which leads to a decrease in the computational time required for classification algorithms and an enhancement in the accuracy of classification.

3.3.1 Spotted Hyena For Effective Feature Selection

Input (K) refers to a preprocessed picture of a spotted hyena population, where $((i = 1,2,\dots n)$.

Output: selected features

```
Set the initial value of the vector parameter
The fitness function is determined for each search space
$K \rightarrow$ best search path
Consider the nearest solution
While (N, num_of iterations) do
    For each search route
        Update the present position
    End for
    Search path threshold check
    Evaluate each search path's fitness function
    Update value if solution > previous search till optimum solution.
End while
Return to search route
End
```

CONVOLUTIONAL GATED ATTENTION RECURRENT NEURAL NETWORK CLASSIFIER

The combination of attention mechanism with Gated Recurrent Units in a RNN to effectively highlight significant cervical cancer features, therefore enhancing the model's ability to learn and identify disease traits. The developed architecture's framework is shown in Figure 2. Initially, a CNN that has been trained on a disease classification task. This means that attributes are stored as features of CNN, which are tensors of feature mappings obtained from a specific yield of a particular convolutional layer. The CNN features may be regarded as additional properties of activations, equipped with filters that have the same number of channels as the convolution layer. The newly introduced characteristics are then divided into smaller sub-parts of equal size in order to get local activations over many areas that include all of the attributes. The newly gained local CNN features may be utilized to sequence and input a GRU-based RNN. This allows for the implementation of an attention mechanism to accurately identify significant segments or elements within the CNN features. The approach improves the active attributes area within each sub-part and info gain over numerous CNN architecture sub-parts. The goal of optimization is to reduce prediction error.

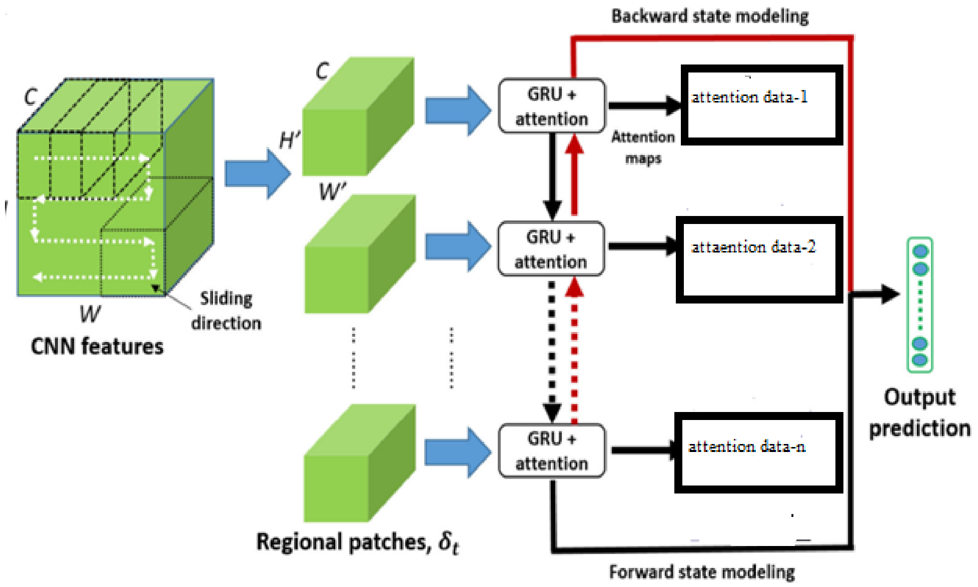


Figure 2: Architecture of Convolutional Gated Attention Recurrent Neural Network

3.4 RNN architecture with attention

The features of cancer qualities are represented as I, whereas CNN convolutional layers provide feature maps as $\delta \in R^{H \times W \times C}$, where H symbolize height, W and C symbolize width, and channel count. The CNN model is pre-trained and supplemented with target classes. δ ,the global feature map is sliced in the gliding path to create T area feature maps $[\delta_1, \delta_2, \dots \delta_T] \in \delta$. In Fig 2, RNN module receives the feature maps. The RNN connects numerous feature maps to learn visual patterns and describe their spatial association. It adds a new step to the cognitive process that is intentional and helps find important areas by giving different parts of attributes different levels of importance for certain traits. The μ_t terms, which are also called the regional attention map, are used to implement the process. These words control the impact of t-th state's pixels and are taught by means of the neural network. The importance increases as the value of μ_t increases. The attention module shows each pixel's proportionate contribution in T regional feature maps. The function of attention is $g: \mu_t, h_{t-1} \rightarrow \varepsilon_t$ in a more formal manner as follows:

$$\propto_t = tanh(\mu_t W_\mu + h_{t-1} W_h) W_a \tag{13}$$

$$\mu_t = softmax\ (\propto_t) \tag{14}$$

$$\forall_t = \boldsymbol{g}(\mu_t, h_{t-1}) = \textstyle\sum_{i,j} \mu_t, ij. \propto_t, ij \tag{15}$$

Where the embedding matrices are denoted as $W_\delta \in R^{C \times C}, W_h \in R^{E \times C}, W_a \in R^{C \times 1}$. The dimensionality of the GRU cell is E, and $\propto_t, ij$ represents the assay of t-th regional feature map located at $(i, j) \in H' \times W'$. It should be noted that ε_t represents the output representation, also known as the feature vector.

3.5 Gated Circulation Unit

The GRU and LSTM operate in a similar manner. However, the GRU unit employs a concealed state to integrate the forgetting and input gates into a unified update gate. The process regulates the extent to which information from the hidden layer of the preceding instant is to be disregarded, as well as the extent to which memory information from the hidden layer of the present moment is included. Additionally, the GRU incorporates a novel "gate" known as the reset gate, which regulates the computation. The GRU network's output representation is expanded as h and thereafter inputted into a fully connected layer in order to facilitate the linear regression layer. The underlying mechanism of the completely linked layer may be expressed as follows:

$$o = f(Wh + b) \tag{16}$$

Where the result generated by the fully connected layer is denoted as o, the transformation matrix is represented as W, the bias is b, and f(.)s is the activation function . The activation function used here is ReLU. The fully connected layer output o is then inputted into a regression layer, resulting in the prediction of the resolution layer for the ith working state. This prediction may be expressed as follows:

$$y_i' = W_{Oi} \tag{17}$$

The categorization of individual cells facilitates precise and expeditious identification of cancerous cells inside the cervix.

PERFORMANCE ANALYSIS

The experimental findings were conducted using Python software, with the analytic parameters including precision , accuracy, F1-score and recall. The parameters pertaining to the developed ConvGateRNN are compared with the existing methods, namely SVM, Decision Tree (DT) and ECDTA Classifier [19].

Implementation setup- The implementation of our pipeline is powered by the Pytorch and Pytorch3d frameworks. Due to the much lesser quantity of genuine training attributes compared to synthetic training attributes, we create a mini-batch consisting of half real attributes and half synthetic attributes. This approach allows us to concentrate on the real dataset. The training procedure is completed after a total of 250 epochs.

- The performance of these category of instances is mostly evaluated based on classification accuracy. The calculation of classification accuracy for two class classifications may be determined using the following equation.

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

- The probability of misdiagnosing an infected individual as healthy is higher compared to inadvertently identifying a healthy person with cervical cancer. The proposed approach is evaluated using precision and recall instead of accuracy. Equations (8) and (9) may be used to compute these performance assessment indicators

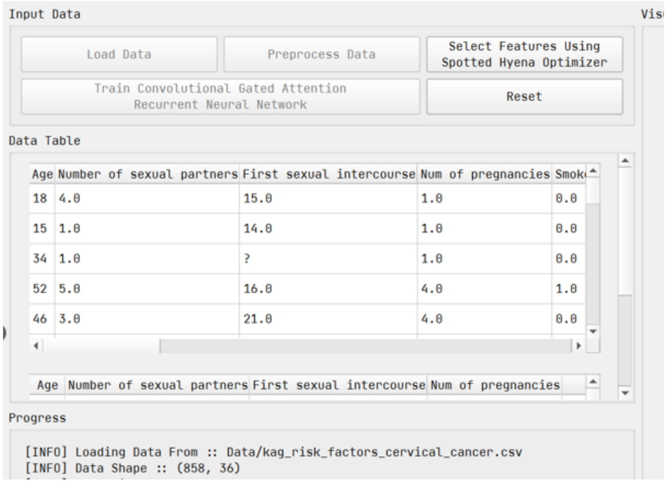
$$precision = \frac{TP}{TP + FP} \times 100$$

$$recall = \frac{TP}{TP + FN} \times 100$$

- F-score: The metric quantifies the proportion of genuine positives to the sum of true positives and false positives.

$$\mathrm{F-score} = \frac{\mathrm{2TP}}{\mathrm{2TP + FP + FN}}$$

As seen in Figure 3, the characteristics that are calculated or loaded via a costly activity such as an API request may be loaded in a lazy manner



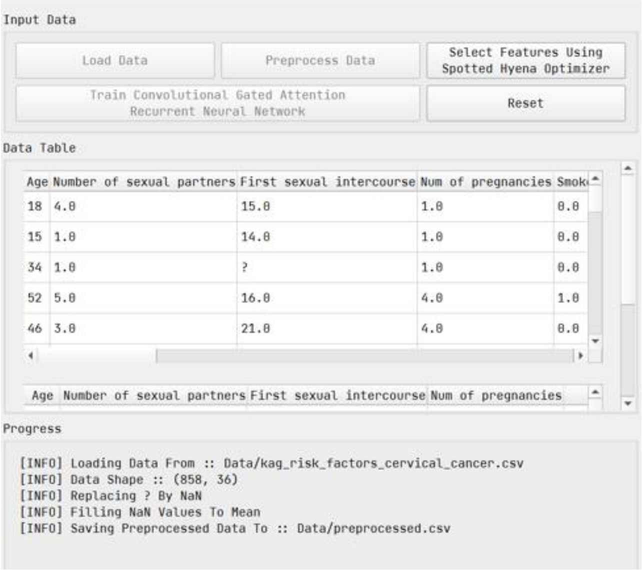
Age	Number of sexual partners	First sexual intercourse	Num of pregnancies	Smok
18	4.0	15.0	1.0	0.0
15	1.0	14.0	1.0	0.0
34	1.0	?	1.0	0.0
52	5.0	16.0	4.0	1.0
46	3.0	21.0	4.0	0.0

Age	Number of sexual partners	First sexual intercourse	Num of pregnancies

```
[INFO] Loading Data From :: Data/kag_risk_factors_cervical_cancer.csv
[INFO] Data Shape :: (858, 36)
```

Figure 3: Loading of data

The significance of data preparation in machine learning projects cannot be underestimated, as seen in Figure 4. It guarantees the quality of data, improves the performance of models, allows for the extraction of features, and provides interoperability with various techniques



Age	Number of sexual partners	First sexual intercourse	Num of pregnancies	Smok
18	4.0	15.0	1.0	0.0
15	1.0	14.0	1.0	0.0
34	1.0	?	1.0	0.0
52	5.0	16.0	4.0	1.0
46	3.0	21.0	4.0	0.0

Age	Number of sexual partners	First sexual intercourse	Num of pregnancies

```
[INFO] Loading Data From :: Data/kag_risk_factors_cervical_cancer.csv
[INFO] Data Shape :: (858, 36)
[INFO] Replacing ? By NaN
[INFO] Filling NaN Values To Mean
[INFO] Saving Preprocessed Data To :: Data/preprocessed.csv
```

Figure 4: Preprocessing of data

As seen in Figure 5. The process of feature selection relies on a particular machine learning algorithm that we want to apply to a given dataset. The proposed methodology employs a greedy search strategy, whereby all potential combinations of characteristics are assessed against a predetermined evaluation criterion.

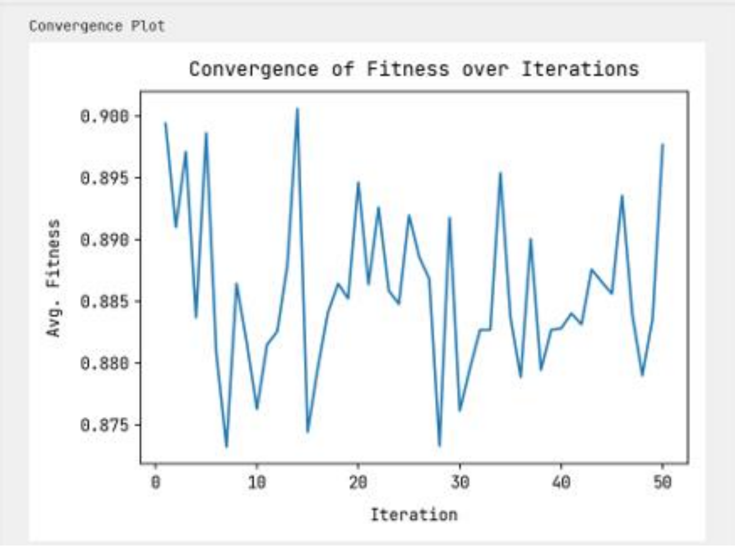


Figure 5: Selection of features

The confusion matrix in Figure 6 depicts the characteristics used for testing and training in the proposed classifier. Rows indicate the anticipated class, while columns provide the key illness prediction data. The crosswise blue hues reflect properly and erroneously classified tested and trained networks. The row below shows each class's implementation, whereas the right column shows each expected class. The confusion matrix illustrates the classification performance of the model in distinguishing between "Normal" and "Abnormal" cases. Out of the total samples, the model correctly classified 838 instances as "Normal," demonstrating strong performance for this class. However, it misclassified 2 "Normal" instances as "Abnormal." On the other hand, the model correctly identified 13 instances as "Abnormal," but it misclassified 5 "Abnormal" cases as "Normal." This indicates that while the model performs well in recognizing "Normal" instances, it struggles with correctly identifying "Abnormal" cases, resulting in a higher rate of false negatives. This disparity may suggest a need for improved feature extraction or addressing class imbalance to enhance the detection of "Abnormal" cases.

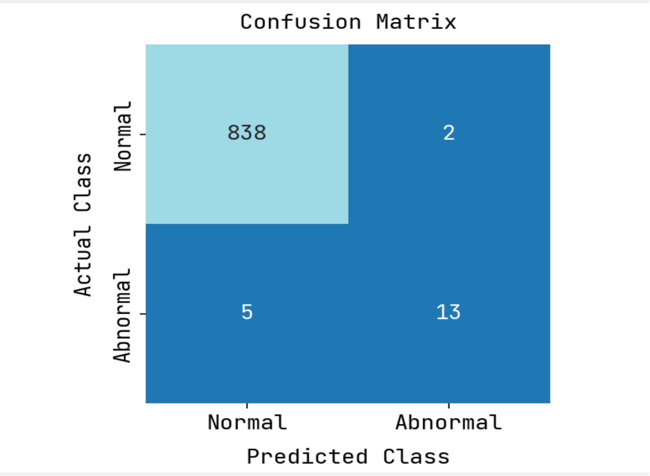


Figure 6: Confusion matrix

The validation and training accuracy meant for ConvGateRNN stands at Figure 7. The graph illustrates the model's training and validation accuracy throughout 100 epochs. To begin with, both training and validation accuracies started at a relatively low value, around 0.6. However, as training progresses, the accuracies improve steadily. By approximately 20 epochs, the model achieves a significant accuracy improvement, exceeding 0.9 for both training and validation. After this point, the accuracy stabilizes and reaches near-perfect levels, with both training and validation accuracies approaching 1.0. The close alignment of the training -validation curve of accuracy indicates that the model generalizes well to concealed data, showing no signs of overfitting or underfitting. This suggests that the model is well-optimized and effectively captures the fundamental trends within the data.

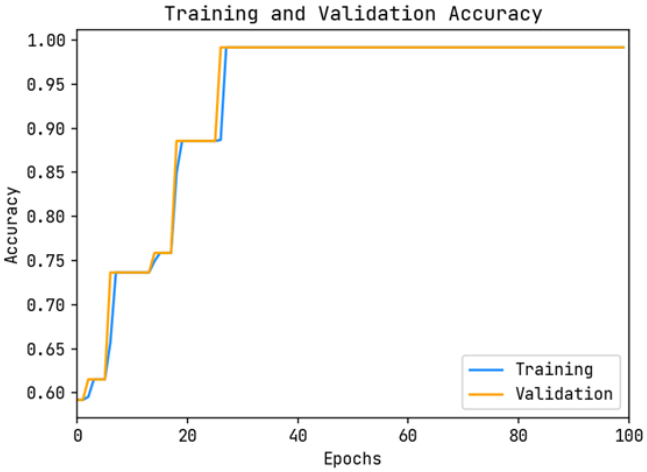


Figure 7: Training and validation accuracy

Loss of ConvGateRNN during the training-validation phase is illustrated in Figure 8 The graph illustrates the model's recital loss during training coupled with validation over the course of 100 epochs. Both loss curves begin at approximately 0.675 and decrease steadily and consistently as training progresses. By the end of the training process, the loss values for both training and validation datasets converge to around 0.650, indicating a reduction in the error rate of the model. The nearly identical behavior of the training together with the validation loss curvatures reflects that the model effectually balances fits the data used during training while takes a broad view to the validation data. This suggests the absence of significant overfitting or underfitting during training. However, the relatively small decrease in loss values over 100 epochs may indicate that the model's learning rate is slow, or the problem may have limited scope for further improvement.

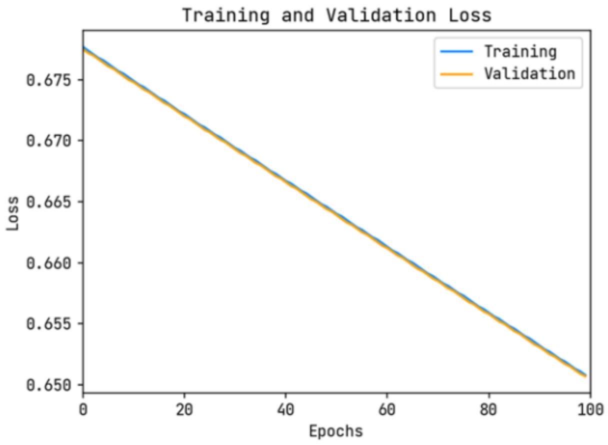


Figure 8: Training and Validation loss

The curve of precision-recall for ConvGateRNN is shown in Figure 9, with the recall value and precision value represented in x-axis and y-axis. During the process, the AP achieves 0.9995 for normal and 0.8325 for abnormal.

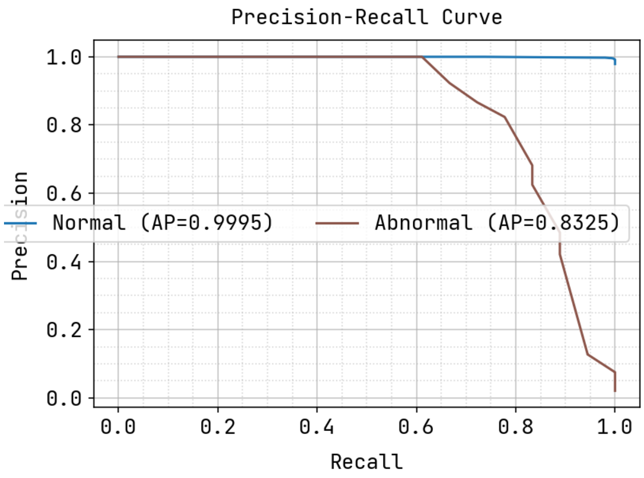


Figure 9: Precision-recall curve

Figure 10 displays a ROC curve, with the x direction has false positive frequency and the y direction has true positive frequency. In this process, the ROC reaches the maximum range of 1 for all categories of cervical cancer, with a threshold range of 1

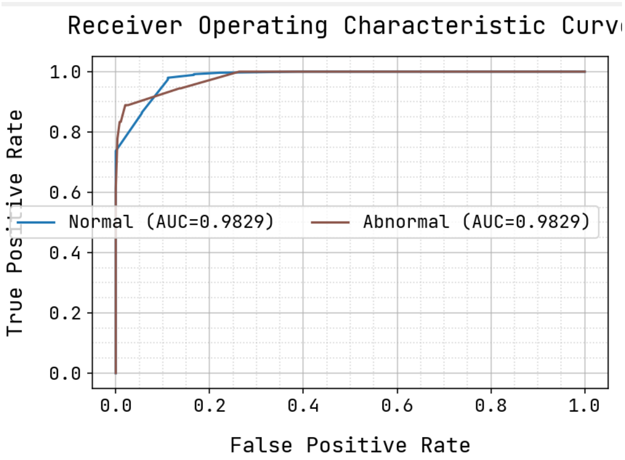


Figure 10: ROC curve

Table 2: Comparative analysis of proposed and existing methods

Parameters	SVM	DT	ECDTA Classifier	ConvGateRNN
Accuracy(%)	81.55	84.54	92.23	96.18
Precision(%)	79.13	86.23	95.54	93.04
Recall(%)	88.34	88.18	88.78	85.99
F1-score (%)	83.28	83.51	92.98	89.19
Log loss(%)	-	-	-	65.07

As shown in figure-11, among the models, ConvGateRNN achieves the highest accuracy of 96.18%, indicating it is the most reliable in correctly classifying instances overall. As shown in Table 2, the ECDTA Classifier follows with 92.23%, while DT and SVM achieve 84.54% and 81.55%, respectively. Regarding precision, which quantifies the ratio of correctly identified positive predictions, the ECDTA Classifier excels with 95.54%, closely followed by ConvGateRNN at 93.04%. DT achieves 86.23%, and SVM performs the lowest at 79.13%. For recall, gages the model's capability to detect all actual positive occurrences, the ECDTA Classifier again performs well at 88.78%, closely followed by DT (88.18%) and SVM (88.34%). ConvGateRNN has the lowest recall at 85.99%, indicating it might miss some positive instances despite its high overall accuracy. The F1score, a metrical that harmonizes precision, recall, highlights the ECDTA Classifier as the most effective with 92.98%. ConvGateRNN achieves 89.19%, and DT and SVM score comparably lower at 83.51% and 83.28%, respectively.

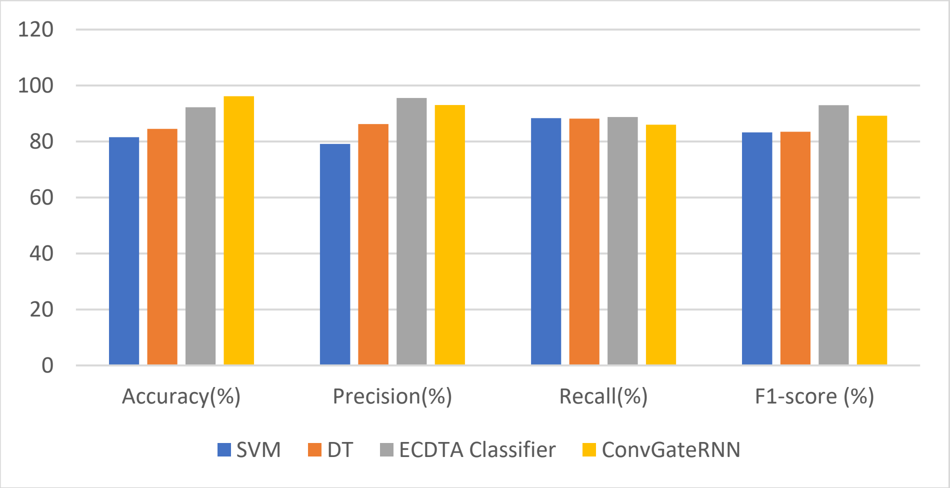


Figure 11: Overall comparative analysis

CONCLUSION

The ConvGateRNN model demonstrates superior performance across multiple evaluation metrics, including accuracy, precision, recall, and F1-score. It achieves an impressive accuracy of 96.18%, significantly outperforming the SVM, Decision Tree and the ECDTA Classifier models. Moreover, the precision, recall, and F1-score of the ConvGateRNN model are notably high, indicating its effectiveness in accurately identifying cervical cancer cases. This suggests that the novel approach of utilizing Convolutional Gated Attention Recurrent Neural Network brings significant advancements in cervical cancer screening.

In conclusion, the ConvGateRNN model provides a promising approach for early detection of cervical conditions, utilizing deep learning methods and attention mechanisms to enhance effectiveness in identifying disease traits. Its superior performance compared to existing methods underscores its potential to revolutionize cervical cancer screening practices, offering accurate and efficient diagnosis at a cost-effective rate. Further research and validation are warranted to fully establish its efficacy and integration into clinical settings.

REFERENCES

[1] Grover A, Pandey D. Anatomy and physiology of cervix. In: Colposcopy of Female Genital Tract. Springer; 2017:3–16.

[2] Payette J, Rachleff J, de Graaf C. Intel and mobile odt cervical cancer screening kaggle competition: cervix type classification using deep learning and attributes classification. Stanford University;2017

[3] Prendiville W, Sankaranarayanan R. Colposcopy and treatment of cervical precancer. International Agency for Research on Cancer, World HealthOrganization;2017

[4] Small W Jr, Bacon MA, Bajaj A, et al. Cervical cancer: a global health crisis cancer. Cancer.2017;123:2404–2412. doi:10.1002/cncr.30667

[5] Jassim G, Obeid A, Al Nasheet HA. Knowledge, attitudes, and practices regarding cervical cancer and screening among women visiting primary health care centres in Bahrain. BMC Public Health.2018;18(1):1–6. doi:10.1186/s12889-018-5023-76

[6] Bruni L, Barrionuevo-Rosas L, Albero G, et al. Human papillomavirus and related diseases report. ICO/IARC Information Centre on HPV and Cancer (HPV Information Centre);2019.

[7] Natae SF, Nigatu DT, Negawo MK, Mengesha WW. Cervical cancer screening uptake and determinant factors among women in Ambo town, Western Oromia, Ethiopia: community-based cross-sectional study. Cancer Med.2021;10(23):8651–8661. doi:10.1002/cam4.43698. World Health Organization.

[8] Arora M, Dhawan S, Singh K. Exploring deep convolution neural networks with transfer learning for transformation zone type prediction in cervical cancer. In: Soft Computing: Theories and Applications. Springer; 2020:1127–1138

[9] M. Suriya, V. Chandran, and M. G. Sumithra, "Enhanced deep convolutional neural network for malarial parasite classification," International Journal of Computers and Applications, pp. 1–10, 2019.

[10] Q. Meng, "Machine learning to predict local recurrence and distant metastasis of cervical cancer after definitive radiotherapy," International Journal of Radiation Oncology • Biology • Physics, vol. 108, no. 3, article e767, 2020.

[11] Chanudom, I., Tharavichitkul, E., & Laosiritaworn, W. (2024). Prediction of Cervical Cancer Patients' Survival Period with Machine Learning Techniques. Healthcare Informatics Research, 30(1), 60.

[12] Aljrees, T. (2024). Improving prediction of cervical cancer using KNN imputer and multi-model ensemble learning. Plos one, 19(1), e0295632.

[13] Glučina, M., Lorencin, A., Anđelić, N., & Lorencin, I. (2023). Cervical cancer diagnostics using machine learning algorithms and class balancing techniques. Applied Sciences, 13(2), 1061.

[14] Devi, S., Gaikwad, S. R., & Harikrishnan, R. (2023). Prediction and detection of cervical malignancy using machine learning models. Asian Pacific journal of cancer prevention: APJCP, 24(4), 1419.

[15] Chauhan, N. K., & Singh, K. (2022). Performance assessment of machine learning classifiers using selective feature approaches for cervical cancer detection. Wireless Personal Communications, 124(3), 2335-2366.

[16] Ali, M. M., Ahmed, K., Bui, F. M., Paul, B. K., Ibrahim, S. M., Quinn, J. M., & Moni, M. A. (2021). Machine learning-based statistical analysis for early stage detection of cervical cancer. Computers in biology and medicine, 139, 104985.

[17] Lilhore, U. K., Poongodi, M., Kaur, A., Simaiya, S., Algarni, A. D., Elmannai, H., ... & Hamdi, M. (2022). Hybrid model for detection of cervical cancer using causal analysis and machine learning techniques. Computational and Mathematical Methods in Medicine, 2022.

[18] Ramzan, Z., Hassan, M. A., Asif, H. M., & Farooq, A. (2021). A machine learning-based self-risk assessment technique for cervical cancer. Current Bioinformatics, 16(2), 315-332.

[19] Anusha R, Dr Srinivas Prasad, A Machine Learning-Based Classifier For Anticipating Risk Factors Associated With Cervical Cancer, Journal of Theoretical and Applied Information Technology, 2023.

[20] https://archive.ics.uci.edu/dataset/383/cervical+cancer+risk+factors