ABSTRACT Rice disease classification is vital during the cultivation of rice crops. However, rice diseases were initially detected by visual examination from agricultural experts. Later the detection process progressed to automation, which involved images. The images captured lead to a lack of supporting information. The traditional approaches are less accurate when used with real-time images. To address this limitation, a novel Rice Transformer is proposed in the paper that merges inputs from agricultural sensors and image data captured from the fields simultaneously. The proposed system consists of two branches: the sensor and image branches. Specifically, the attention approach is employed to extract the features from both modalities. Later, the extracted features are sent to the cross-attention module as input in a crisscross fashion, enhancing the ability to identify the features specific to rice diseases. The extracted features are further pooled, merged, and later passed through the Softmax classifier to classify the rice disease precisely. The dataset collected is a customized dataset with 4200 samples collected on a real-time basis from rice farms. The experiments conducted on the dataset represent that the proposed approach outperforms all the other fusion and attention models considered for comparison in this paper. The ablation analysis and performance metrics are measured to determine the effectiveness of the proposed system. The results achieved are quite promising as the proposed Rice transformer model achieves an accuracy of 97.38% for controlling rice disease.

INDEX TERMS Rice transformer, rice diseases, multimodal, data fusion, self attention, cross attention.

I. INTRODUCTION

The yield of agricultural production is significantly affected by crop disease outbreaks. A large outbreak of a disease can destroy crops that have been extremely difficult to grow, resulting in irreversible harm [1]. Along with large outbreaks, even the emergence of small-scale diseases can seriously affect crop yields and quality. Therefore, establishing the correct classification of leaf diseases in crops is essential [2]. Rice is an essential food for people all over the world. Hence large population is dependent on rice for their food needs. The most common diseases in rice are blast, brown spot, and blight [3]. If infections are not treated on time, it can lead to huge economic and yield production loss. Therefore, many agricultural researchers are devoted to studying the methods for detecting rice diseases that can further assist the farmers in making decisions precisely [4]. As far as crop diseases are concerned, researchers have made some remarkable advancements in the classification of crops. However, the most existing approaches for diagnosing and classifying rice diseases are by using images and applying image processing approaches.

Nevertheless, the traditional image processing approach used hand-crafted features, and thus, performance could not reach the mark. The images input also emerged over time, starting with images captured in a controlled environment [5] then moving towards capturing real-time images in the farm [6], hyperspectral images [7] and latest by using Unmanned Aerial Vehicle [8]. Research in crop disease has gradually embraced deep learning technologies [9]. Despite this, Deep Learning (DL) still pose many problems [10]. The traditional CNN, which consists of multiple layers of
activation and convolutional operations, understands attributes from the input images and trains the model according to the learned attributes [11].

Traditional expert diagnosis of rice leaf disease is expensive and vulnerable to subjective error [12]. Artificial Intelligence [13], machine learning, computer vision, the Internet Of Things [14], and deep learning [15] are commonly utilized in crop disease diagnosis due to the rapid advancement of computer technology [16]. Traditional machine vision algorithms use color, texture, and shape features to segment RGB images of crop diseases. The crop diseases are similar in many ways, it can be challenging to identify the type of illness, and disease recognition accuracy can be low in natural conditions. Unlike traditional learning processes, CNNs do not require complicated image preprocessing and feature extraction and instead offer an end-to-end solution that substantially simplifies recognition [17], [18] Utilizing CNN for automatic disease identification improves diagnostic accuracy in real-life farming situations and reduces labor costs [19], [20].

Agriculture research is enriched with a variety of data that can be obtained from a variety of sources including IoT sensors, vegetation indices, images from unmanned aerial vehicles (UAVs), and satellite imagery. In order to understand crop growth circumstances and disease symptoms, fusion techniques must integrate multiple forms of retrieved data. Data fusion using machine learning has also advanced significantly. When applied to agriculture data, it will significantly impact disease diagnosis as well as plant protection. In Precision Agriculture, agricultural data is combined with artificial intelligence algorithms and fusion algorithms to develop crop growth monitoring and crop protection tools. Using AI techniques and images collected from the field, this paper aims to solve the problem of automatic disease diagnosis of rice. Diagnoses of diseases require robust algorithms that can deal with the challenges present in the process.

Further, unusual circumstances, such as significant light variation or background clutter, can pose severe problems in raw images. It may not be possible to use low-level visual characteristics in this case. The Internet of Things (IoT) plays a key role in this scenario, as it offers various ways to gather precise soil data that can be retrieved as very relevant characteristics in aiding modern identification methods. The accuracy and robustness of multimodal applications is higher than that of single modalities. They integrate information from a variety of sources at a signal- or semantic level. A major advantage of applications with multiple modalities is that they perform better than applications with one modality. This is due to the advancement of deep learning techniques, increased computing infrastructure, and large datasets [21]. The study in [22] represents that a model deployed with multiple modalities perform better over single modality.

In deep neural networks, the Attention Mechanism attempts to emulate the cognitive process of focusing on certain relevant information while disregarding non-relevant information. The relationship among the modalities can be represented by attention mechanism [23]. Fusion approaches underperforms to manage the missing modalities hence attention frameworks are utilized to manage the missing and noisy data [24]. The attention network has become widely employed in machine translation, and other fields in recent years due to its ability to extract discriminative properties of the area of interest. However, in the domain of agricultural disease classification, attention approach is still in the exploratory phase.

Researchers in [25] enhanced the identification rate of grape illnesses in the Plant Village dataset to 99.14 % by adding the attention module with backbone architecture as ShuffleNet. There are various types of attention mechanisms such as self attention wherein temporal relationship amongst the same modality is captured. Hierarchical attention is transferring information from lower level to higher level [26]. A particular aspect of information is given attention, such as certain features, regions in an image, or a specific moment in the sequence. In some instances, background features overwhelm leaf and diseased regions in the foreground, resulting in poor model performance on various test images. By incorporating an attention mechanism to CNN allows the model to focus on those features that are meaningful instead of global features. Several researchers have adapted the attention mechanism for the classification of plant diseases after it demonstrated promising results on a number of image classification datasets. Recent studies have shown further improvements by applying deep learning algorithms. Hence, there is a paradigm shift in the research area that focuses on the amalgamation of multimodal and attention approaches.

Therefore, motivated with the idea of merging multimodal and attention approaches, this paper proposes a novel Rice Transformer model that automatically extracts important rice disease features from a complex environment. The important disease related features are extracted, and the irrelevant features are discarded. The architecture of proposed Rice Transformer is represented in Figure 5. Using an improved CNN network model, this paper proposes a reliable way to diagnose rice leaf diseases.

In summary the contributions of this paper are enlisted as follows:

- A novel multi stream framework based on multimodal data input which can extract deep features is developed.
- A model that is robust and accurate is developed when compared with the state of the art methods is represented by performing extensive experimentation and ablation analysis.
- A multi branch self attention encoder is used to extract the features from the agricultural sensors and image data.
- A novel cross attention module is devised that learns the variance and correlation amongst the different diseases by sensor attention module and image attention module and integrates the extracted attention maps via cross fusion mode to gain the relevant features of particular rice disease.
The further skeleton of this paper is structured as follows. Section 2 reviews the literature of the existing systems followed by section 3 which briefly introduces the concepts that are used in the paper. Section 4 represents the experimental analysis carried out. Section 5 presents the experimental analysis and Section 6 concludes the paper along with the futuristic scope of the model.

II. RELATED WORK
Although fusion technique is a rapidly spanning area in the research based on agriculture, there are very few studies on disease classification with this technology. The literature presents a variety of applications of data fusion to agriculture, including disease detection, classification, crop identification, yield prediction, and land monitoring. However, the section below focuses on the literature on disease diagnosis in plants.

A. STATE OF THE ART APPROACHES IN THE IDENTIFICATION OF CROP DISEASES
Initially, image processing was employed to classify, diagnose, and predict diseases in crops. The features such as color, shape, and texture are extracted to obtain the region of interest [27], [28]. Over the past decade, the deep learning approach has been considered more reliable due to its capability of learning visual features. Moreover, transfer learning is a collection of fine-tuned methods that allow for the development of high precision frameworks on more constrained specialized datasets, such as those pertaining to plant diseases [29]. It is perceived that the fine-tuning approach is superior to a CNN model that is trained from scratch. The ANN approach became popular in the agricultural domain and thus was applied to the crop disease detection domain [30]. On a similar basis, Faster RCNN was used to control the rice diseases that overcame the limitation of overfitting. YoloV3 model was used to scan the rice farm video to detect the exact location of the illness on the leaves [31]. IoT based agriculture enhances the quality of agricultural production all over the world. In [32] SMS alert system is proposed that alerts the farmers regarding the emergence of diseases on crops. Classification is done using SVM and K-means. A pesticide spraying plan is also suggested based on the prediction. The alert system assists the farmers with supporting information.

B. DATA FUSION IN AGRICULTURE
There is a tremendous amount of data generated by the tools used for diagnosing plant diseases and classifying them. There are two options to deal with such type of data. The first option would be analyzing individual modalities, comparing the results produced, and assessing the strength of each modality. The second option is merging the modalities to diagnose the condition of crop diseases [33]. The very first attempt to detect diseases using multisource data was to join meteorological and satellite data. A logistic regression model was used to classify and obtain the influential factors from both modalities. The results indicated that multimodal data integration has great potential for disease detection. In [34], the authors propose a multi-context fusion network for detecting crop diseases. The approach was based on field images and contextual data such as longitude-latitude information, season, and climatic parameters. A model comprises three components: CNN for feature extraction from images, ContextNet for making contextual features fusions, and a fully connected network for combining all features. The accuracy achieved is 97.5%. The dataset collected has variance because of climatic and geographic difficulties in dealing with different crop diseases. The researchers split another study into three phases to detect diseases in banana fruit using object-based localization, pixel-based classification, and diagnosis of diseases in banana fruit in African landscapes. SVM was used as the classification method, along with multispectral bands, vegetation indicators, and UAV images [35]. An object detection model named RetinaNet was developed to detect banana plants in the field as well as to detect diseases in banana plants based on UAV images captured in the field. The custom classifier provided better results than the VGG model, achieving a 92% accuracy. This method has good classification results; however, it incurs considerable training time. In [36], a Rice Fusion architecture is proposed that merges data from two modalities as input and classifies three categories of rice diseases. The proposed approach is compared against various fusion approaches; however Rice Fusion model outperforms all other models by obtaining an accuracy of 93.51%.

C. ATTENTION APPROACHES IN AGRICULTURE
The authors in [6] propose an architecture that extracts a feature map from an infected image of a plant trained CNN from a photo of a contaminated plant. They are further divided into multiple patches that slide in a “snaking” pattern. The patches are then fed into Gated Recurrent Units, which bi-directionally share relevant information to update an internal representation of plant disease. To infer discriminative and local features, the soft attention mechanism is used. The authors in [37] have focused on tomatoes, potatoes, and corn as these are the most cultivated plants in China. The authors in [38] have used a unique Convolutional Block Attention Module (CBAM) to improve the classification of the diseases. The system achieves an overall accuracy of 99.5%. The paper employs various attention mechanisms to classify tomato diseases. Self-attention, Dual attention, CBAM, and Self Excitation attention module are compared. The results show that CBAM based approach outperforms the other three attention models for classifying tomato diseases by achieving an accuracy of 99.69%. In [39], a self-attention mechanism based on location is employed on the baseline of BaseNet architecture to identify minor lesions on rice and cucumber leaves. The approach compares three variants of attention mechanism: the concatenation of two Self-attention modules, one SA module, and the third is parallel addition of two SA.
However, the approach with one SA outperforms the others and reaches the accuracy of 95.33%.

After reviewing the literature, it is observed that automatic plant disease diagnosis, classification, estimation is deployed using traditional approaches such as image processing, applying machine learning and deep learning classifiers to the data obtained, usage of sensors. They later shifted the paradigm to the data fusion approaches, which proved its efficiency in Precision agriculture. However, the fusion approaches had the constraint of obtaining massive real-time data, which is not always possible. Therefore, to overcome the limitations of the current method for controlling plant diseases, attention mechanisms have been developed that enable more accurate predictions, diagnosis, grading, and recommendations.

III. THEORETICAL BACKGROUND
A system is said to be reliable and robust when information from various data sources is fused. As a precursor to the proposed method, concepts such as attention techniques and multimodal information fusion approaches based on the Artificial Intelligence domain need to be understood and are available in the literature. This section contains a brief discussion of these methods.

A. MULTIMODALITY
Modality describes how things are perceived through sensory inputs such as hearing, sight, smell, taste, and touch. Modality refers to how information is depicted and conveyed among people. Multimodality could be seen in audio, language, physical sensor signals, physiological signals, and vision. Humans can describe objects or phenomena using multimodal data by combining different aspects with complementary or auxiliary information. Models with multiple modalities outperform single modality models.

B. INFORMATION FUSION
The data collected from various input sources are jointly used to perform a task. The features from different modalities either supplement each other to boost the system’s performance or complement each other. Multimodal data fusion is categorized into late fusion (decision), early fusion (feature), and hybrid fusion (intermediate) based on the scale of the model at which categorizations are merged. The fusion approach is task specific, domain specific, and data specific. There is no precise set of rules for fusion approaches. The intra modality differences will not be captured using the early fusion approach, and the intermodality discrepancies will not be captured using the late fusion approach. So as per the requirement of the model, the fusion approach is chosen and further used for deployment. Concatenation, multiplication, or weighted sum are the various methods to join the modalities. Figure 1. represents the concept of information fusion for the proposed model.

C. MULTIMODAL CROSS ATTENTION MODEL
The inputs to the model come from different data modalities. The information from two distinct modalities is fused and given input to the model. In the multimodal cross attention model, the inputs are fed to the encoder linearly. Later, the output from the encoders is not directly inputted to the decoders but is provided crossly. The features from both modalities that complement each other are multiplied and later concatenated to predict the desired output. Figure 2. shows the concept of the multimodal cross attention model.

D. TRANSFORMERS
The Transformer is a black box model in deep learning that takes input and produces output. Inside the Transformer, there are two components, namely Encoder and Decoder. The input information is given to the encoder, and the decoder produces the result. The encoder and decoder block further consist of multiple encoders and decoders. The count of the encoders can differ from model to model. The encoder count that gives good results can be considered in implementation. The number of decoders is the same as the number of encoders. The input is given to the first encoder. The output from the final encoder is provided as input to each decoder in the decoder block. Figure 3. represents the structure of the Transformer.

According to Figure 1, the data collected from agricultural sensors is in continuous form, and therefore Multi Layer Perceptron architecture is the best choice to extract
features from this modality. CNN architecture has proven very effective in extracting features from the rice images. Both modalities have distinct properties and are not related to one another in terms of time. As a result, based on agricultural features derived from the MLP architecture and image features fetched from the CNN architecture, a novel framework, an amalgamation of multimodal fusion approach and attention technique, is proposed to control diseases in rice crops.

IV. MATERIALS AND METHODOLOGY: RICE TRANSFORMER

This section discusses about the data collection strategy along with the working of proposed Rice Transformer architecture.

A. IMPLEMENTATION SPECIFICATIONS

The recommendation of agricultural advisory based on the diseases affected by the rice crop is a classification problem since the model uses agricultural sensors data and image data to describe the rice disease categories. The diseases are first classified as healthy, blight, brown spot, and blast. Based on the input features, the model predicts the rice disease class. The multimodal data for the proposed model is gathered in two steps: first, the readings from the agricultural sensors deployed on the rice farms, and second, the rice leaf images collected simultaneously from the rice field location. The input is collated from three different sensors. DHT22 sensor captures Relative Humidity (RH) and Temperature (T), the soil moisture resistive sensor captures soil moisture (M) readings from the soil, and lastly, the analog soil pH sensor (NPK) collects the pH value of the soil in the field. The values from all the sensors are continuous in nature. Four thousand two hundred samples for all four parameters along with the simultaneously clicked rice image are provided as input. The number of sensor data samples and images samples are equal as the data collated is simultaneous. In this paper, three rice disease categories and one healthy class are selected for categorization because these categories are commonly found diseases in all the growing stages of rice crops. Implementing an integrated management system enables the proposed cross attention framework to adapt to diverse geographical locations. The images dataset consists of 4200 images of four different categories of rice diseases with 840 samples in each class of rice disease classification. The sample collection is followed by dataset division into training and testing datasets in an 80-20% ratio. A significant proportion of the image data were captured with a Charged Coupled Device (CCD) from October 2018 to November 2021. In addition, the rice images are augmented randomly by flipping the images horizontally and vertically so that the diversity of the training images increases. The photos are annotated using an open-source tool named Make Sense. The agricultural experts authenticated the image annotations from different Research Extension Centers. The data from agricultural sensors is combined with the data acquired from the rice images and is used in the training and testing stages of the afresh-designed multimodal cross attention architecture. To assess the classification performance of rice diseases, parameters such as Accuracy, F1-score, Recall, Precision, and Confusion Matrix are employed.

B. DATA COLLECTION AND PREPROCESSING

To the best of the authors’ knowledge, no data consisting of real time images and agro meteorological sensors for the representation of rice disease has yet been collected and available in the open domain for direct use. Hence, in this work, data of the sensors and images is collected manually for model training and validation purposes. The experimental data is collected through an array of three agro-meteorological sensors as well as using CCD Camera. The sensors were placed in the rice field and simultaneously images of the rice plants in the same field were clicked. Sensor readings and images are recorded for each of the four rice health categories were collected for approximately two years. A data set in total consists of 4200 samples where each class has 1050 samples. A total of 4200 sensor inputs and images and their corresponding 4200 labels are considered in this experimentation. A train-test of split of 80:20 was done such that out of total images, 3360 are used for training and 840 images are used for testing purposes. A few representative samples for four classes with the image and corresponding agro-meteorological array data are shown in Table 1. The images in the table represents the rice infection classes i.e. brown spot, bacterial blight, rice blast and healthy moreover the corresponding agro-meteorological parameters i.e. temperature value, humidity value, soil moisture value and pH value are shown in the last row of the table respectively.

C. ARCHITECTURE

This paper proposes a novel multimodal architecture named Rice Transformer for controlling rice diseases. The proposed model uses two types of information: agro-meteorological data obtained from the sensors and rice image information obtained from the rice fields to achieve the best classification performance. Figure 5 represents the Rice Transformer architecture. As depicted in Figure 5, the Rice Transformer architecture comprises two branches: agrometeorological and rice image streams. The sensor’s agro-meteorological readings of various parameters are passed through Multi-Layer
Perceptron architecture to extract high-level features from the raw agro-meteorological data. Similarly, the rice images are passed through the Convolutional Neural Network layer to extract high-level semantic representations. As seen in Figure 5, the extracted high-level features from both the branches are then given as input to their respective self-attention encoder. The query, key, and value vectors generated from the self attention modules are given as input in a criss-cross fashion to the cross-attention decoder module. This aids in discovering interactive data within agro-meteorological and image feature sequences. The agro-meteorological cross attention stream and rice image cross attention stream aids in selecting traits that are more relevant and useful in disease control. Figure 4 represents the training and testing process of the proposed Rice transformer model. For training and testing of Rice transformer model, data is divided into training and testing sets with a ratio of 80:20%. Training data constitutes 80% of the data. The remaining 20% is for testing the model. To begin with, CNN and MLP models were created, but only their features were extracted; no final rice disease classification was carried out. MLP and CNN each generate their own outputs, and these are the features generated by them individually. These features are send to encoders and decoders in crisscross fashion. They are then combined with concatenate(). The merged output is then provided to the final layer set. The output dimension of the MLP network is 4-7, and 200-100-50-25 is the output dimension of CNN. Feature vectors from MLP and CNN architectures are fed into the new Keras model. From MLP, a 7*1 vector is extracted, while from CNN, a 25*1 vector is extracted. Concatenating the outputs results in a combined 1-dimensional vector of 32. Two additional layers are then applied. Finally, the Rice Transformer is compiled. 0.001 is the learning rate. Using the Adam optimizer and cross-entropy loss function, the model is compiled. Fit() is used to train Rice Transformer and evaluate its performance on a training dataset. The experiment is conducted for 500 epochs, and a batch size of 32 is used. In order to fine-tune all of the weights, backpropagation is used. After this, one more fully connected layer is added, equal to the number of rice diseases the model is diagnosing. The last layer will have four neurons as the final output classes “Healthy”, “Bacterial Blight”, “BrownSpot”, “Rice Blast”. Rice infection classes are classified using the Softmax function. A crop advisory is generated, including preventive and corrective measures as per the prediction. The following section delves into these blocks in further depth.

**D. FEATURE EXTRACTION BLOCK**

The Rice Transformer model accepts inputs from two different modalities. Therefore, the features from the two input modalities are extracted using two different filter blocks. Multi-Layer Perceptron is used to extract features from agricultural sensors, and Convolutional Neural Network extracts features from rice images.

1) **FEATURE EXTRACTION FROM AGRO-METEOROLOGICAL SENSOR DATA**

The model comprises an input layer, hidden layers, and output layer. While designing the model, multiple weights and bias values are involved. Due to this, the model tends to overfit, which means it tries to fit the training data perfectly. However, if the model is tested with actual test data, it fails catastrophically. The dropout layer is used as a regularization method to solve the overfitting issue. In dropout, the subset of features will be used to improve the accuracy of the model. A dropout rate of 0.2 is applied. A random selection of features will be made based on the dropout rate, and these features will be deactivated. The rest features are forwarded to the subsequent hidden layers and ultimately to the output layer. The backpropagation is used to correct the loss; their weights will be activated, whichever neurons are activated. When testing the model, all neurons will be connected. The product of the weight of all neurons and the dropout ratio is calculated and thus reduces variance. The training dataset comprises readings from agro-meteorological sensors. The input features are multiplied by weights. The summation of bias with the product of input weights and their respective weights is calculated. The summation value is then passed through an activation function, and the output is predicted. The difference between the actual output in the labeled dataset and the output predicted by the model is called loss. The loss value should be minimum, so the weights must be adjusted in such a way that

---

**TABLE 1.** Data samples for rice images and their corresponding agro-meteorological data array obtained for 4 infection classes.

| Rice disease class | Brown Spot | Bacterial Blight | Rice Blast | Healthy |
|--------------------|------------|------------------|------------|---------|
| Sample Image       | ![Sample Image](image) | ![Sample Image](image) | ![Sample Image](image) | ![Sample Image](image) |
| Agro-meteorological sensors measurement (T °C, RH -%, M - analog output, NPK - pH) | 25, 87, 732, 3.2 | 27, 72, 666, 5.4 | 33,93, 810, 4 | 30, 63, 432, 6 |
the predicted output should be similar to that of that actual output. During the training, the weights need to be updated so that the loss value must be decreased. Optimizers decrease the loss value. Thus, backpropagation is applied. The weights are updated by using the learning rate and the derivative of the loss function concerning the specific weight. The learning rate is 0.001 as it achieves a global minima point. This continues for 500 epochs as the model learns the ideal weight at 500 epochs. The model automatically learns the similarity between output and input. The proposed model classifies four rice diseases using the categorical cross-entropy loss function. Adam’s optimization algorithm is used for minimizing the loss. Following compilation, the model is trained, and the model’s performance is assessed using an accuracy performance measure. A one dimensional array of features with dimension $128 \times 1$ is the output from the filter block as a feature vector. Figure 6 shows the feature extraction process of sensor data.

2) FEATURE EXTRACTION FROM RICE IMAGE DATA
Images captured from rice fields are the second input modality to the model. The extraction of most relevant features from the images is necessary. The images are collection of multiple pixels. Pixel is a single smallest element of an image. The existing study related to the research problem concludes that Convolutional Neural Network has proved its efficiency to solve problems that involve images. CNN comprises a feature extraction module, followed by the classification module. In the proposed model, CNN extracts various features from the rice images. The Rice Transformer model performs a sequence of operations such as convolution, padding, pooling, and flattening to extract features. The details of the CNN operations are as follows:

**Convolution:** Images are depicted in numbers. CNN is trained on rice images to recognize the disease pattern and then classify the disease spots on the image. The patterns are known as filters. CNN utilizes filters to identify the essential features of rice images for disease predictions. The filter is a small matrix of weights. The convolution operation is performed between the input image and the filter to produce a feature map. The dimension of the filter is smaller than that of input image. The convolution operation is the element-wise multiplication of the filter and the image input data. The mathematical representation of convolution operation is represented in Equation 1. The elementwise multiplication result is added and produces a single feature map output pixel.

$$Feature \ map = Image \ input \ast Filter \quad (1)$$
To obtain all the elements in the feature map the filter needs to be spanned over the entire input image. The dimension with which this spanning matrix spans over the image is called stride. The spanning matrix will produce a feature map of reduced size. The size of feature map is calculated using Equation 2.

$$
\text{sizeof (Feature map)} = [\text{sizeof (original input)} - \text{sizeof (filter)}] + 1 \tag{2}
$$

The larger the spanning matrix dimension, the smaller the feature map’s size.

The default Activation function Rectified Linear Unit (ReLU) is applied after receiving the feature map that converts the negative values in the feature map to 0 and keeps all the positive values in the feature map unchanged.

Padding: After applying filter, there is some loss in information as the feature map size is reduced. Suppose the output dimension is required to be same as that of the original image input dimension. In such a case, padding is applied so comparatively less information is lost and extracts more specific features. The mathematical representation of padding output is represented in Equation 3.

$$
\text{Padding output} = \text{sizeof (original input)} + 2(\text{padding value}) - \text{sizeof (filter)} + 1 \tag{3}
$$

This operation is repeated a number of times based on the features that need to be extracted.

Pooling: The pooling operation is used after a convolution operation to reduce the dimensionality of the feature map. If the dimensions are reduced, the training time is reduced and prevents the model from overfitting. By applying the pooling operation, the height and weight of the feature map are downsampled; however, the width is preserved. The max-pooling window slides over the input and fetches out the maximum value from the widow and thus reduces the dimensionality.

Flattening: The multi-dimensional vector obtained from the pooling layer is the feature map. For classification purposes, the necessity is providing one-dimensional input. So multi-dimensional vector needs to be converted into a one-dimensional vector, which is done by flattening operation. This 1D vector is forwarded to the cross attention module to predict rice diseases further.

Building of Rice image feature extraction model: The input image is a row, column, and depth multidimensional matrix. The build () method, which requires four parameters: rows, columns, depth, and classes, is used to create a CNN model at first. Because the layers are added serially to the model, it is defined as Sequential. The layers in CNN are added after the initialization of the model. The first convolution layer retains Thirty-two characteristics. By using padding bits, the dimensions of the input image are preserved. A total of two convolution layers were used. The third
convolutional layer learns a total of 128 features. ReLu activation mechanism is applied. To avoid overfitting, the model uses a dropout value of 0.2. On the prior max-pooling layer output, the flattening layer is applied. This becomes a 1-dimensional array after flattening, and 128 characteristics are extracted. The feature extraction process from rice images is represented in 7 A library in deep learning named Keras extracts characteristics from rice pictures. The dataset is split into training (70%), validation (20%), and testing datasets (10%). The model is trained with the fit () function. The model is trained with the Adam optimizer. The loss function utilized is categorical cross-entropy because the proposed approach is a multi-class classification problem. While training the model, it is iterated for 500 epochs with a 0.01 learning rate and 32 batch size.

**E. SELF ATTENTION ENCODER BLOCK**

To calculate self-attention, initially, the feature vectors generated out of agro meteorological sensor feature extraction block and image feature extraction block with dimension \(128 \times 1\) are provided as input vectors to the agro-meteorological and image encoder, respectively, to produce a triad of vectors. Three vectors, namely Query, Key, and Value vectors, are created for each input. To generate these vectors, at first three weights, \(W_q\), \(W_k\), \(W_v\), are created, and its value is initialized randomly. The value of these weights is changed based on Backpropagation. The loss value is calculated and updated in the Backpropagation. The input is multiplied with all three weights, and the respective \(Q(i)\), \(K(i)\), and \(V(i)\) vectors are obtained as shown in Equations 4, 5, and 6 respectively. The dimensionality of the input vector is 64 dim. The dimensionality of vector triplet is kept as 64 as this dimension is a hyperparameter. Better results were obtained when the dimension was set to 64. The dimension values are hyper-parameterized.

\[
Q(i) = \text{Input}(i) \ast W_q \quad (4) \\
K(i) = \text{Input}(i) \ast W_k \quad (5) \\
V(i) = \text{Input}(i) \ast W_v \quad (6)
\]

After finding the queries, keys, and values, the next step is to calculate the attention score. The score determines how much emphasis to give to the features. The attention score is computed by taking the scalar product of the query vector and the associated key vector for all the inputs. The mathematical representation to calculate attention score is represented in Equation 7.

\[
\text{Attention Score} = Q(i) \ast K(i) \quad (7)
\]

The next step is to divide the attention score by the square root value of the key vector dimension. The key vector dimension used is 64, so the square root value becomes 8. By using 64 as hyperparameter value of key stable gradients is achieved. The scores are passed through a softmax activation function. The Softmax activation function normalizes the scores. It gives the probabilities of the output classes, and all the probabilities add up to one. The class with the highest Softmax probability is predicted as the rice disease. The next step is to multiply the Softmax probabilities with a value vector that gives a ‘Sum’ vector. The Sum vector is mathematically represented using Equation 8. The sum vector will focus on the relevant features and drop down the irrelevant ones. The Sum vector, a weighted value vector, generates the output, which is the output of self-attention layers. This resultant Sum vector is further fed to feed forward neural networks. In real-time the actual calculation is performed in matrix form for fast processing.

\[
\text{Sum} = \text{Softmax} \left( \frac{Q \ast K}{\sqrt{dk}} \right) \quad (8)
\]

The multi-head attention uses multiple such sets of \(Q\), \(K\), and \(V\) weight matrices. The proposed model uses three sets of Encoder and Decoder. Each of these sets is initialized randomly. A similar process of finding out the sum vector is followed for all the three encoders. As there are three attention heads used, three different weight matrices are initialized randomly, and ultimately, three sum matrices are obtained. However, the input to the Feed Forward layer expects a single matrix. So, the three sum vectors generated from 3 attention heads are concatenated and multiplied by an additional weight matrix \(W_o\) that was trained jointly with the model.

The final result is a Sum vector that captures information from all the three attention heads and is now sent as input to Feed Forward Neural Network.

**F. CROSS ATTENTION DECODER BLOCK**

The functioning of the decoder block is similar to that of encoder. Only the difference is that the key and value vectors are passed to the decoder and the rest operation are performed as self attention block. The cross-modal attention module in the agro-meteorological stream accepts the output from the third image encoder as key and value vectors and output from the third encoder of image encoder as key and its own query vector. Similarly, the key and value vectors from sensor encoder is given as input to image decoder along with the query matrix generated by the image self attention layer. Now both decoder have query, key and vector matrices but in criss-cross manner. The final sum score is calculated in the similar manner as it has been calculated in self-attention encoder section. Figure 8 represents the cross attention mechanism in the Rice Transformer. The output from the decoder is then passed on to pooling layer.

**G. POOLING LAYER**

In pooling operation, the inputs are down sampled, i.e., the feature map’s dimensionality is reduced. When the size of the feature map has been reduced, the speed of the model training becomes faster. There are various types of pooling such as max pooling, average pooling, sum pooling, statistical pooling. The proposed model uses max pooling for the image encoder, and statistical pooling is used for the
agro-meteorological encoder. The real time rice images that are captured have their unique features, such as the entire leaf being green in color; however, the diseased portion is brown or yellow also the shape and size of the diseased part varies. It is not always possible that the captured image will always maintain the same capturing angle; it can be distorted and compressed. The diseased portion can be anywhere on the leaf, and the background texture is different. The illumination while capturing the image is not consistent for all images. Even with such differences in images, the CNN model should classify rice diseases accurately because of the unique features of rice leaf images. The CNN model should not look for the features at a specific location, shape, or size because it could not recognize the diseases. Hence CNN model should incorporate the property named spatial invariance, which gives the model the flexibility to recognize an image feature even if the object in the image has different illumination, background color, or orientation. In statistical pooling for E1, the mean vector is concatenated with the standard deviation vector over agro-meteorological features.

For second stream the pooling approach used is max pooling which fetches out the maximum value when the max pool filter slides over the input. The most relevant features are preserved in max pooling as the sharpest features will have maximum value, and the features with the largest value have the feature’s similarities.

\[
Pooled_{E1} = \text{concatenation} \left[ \text{mean} \left( output_{CAM1} \right), \times \text{std deviation} \left( output_{CAM1} \right) \right] \quad (9)
\]

\[
Pooled_{E2} = \text{max pool} \left( output_{CAM2} \right) \quad (10)
\]

\[
\text{Pooling} = Pooled_{E1} + Pooled_{E2} \quad (11)
\]

where output_cam1 is the output generated from Sensor Image(SI) cross attention module and output_CAM2 is the
output generated from Image Sensor(IS) cross attention module. Pooled_E1 is the pooled feature vector of the CAM1 output, and Pooled_E2 is the pooled feature vector of the CAM2 output. The dimension of Pooled_E1 and Pooled_E2 are the same. The pooled feature maps are finally concatenated for rice disease classification and recommend the advisory as per the classification.

**H. CLASSIFICATION**

This is the final layer in the Rice Transformer model that classifies rice diseases into various categories. As the proposed model is a multiclass classification problem, a softmax activation function is used for classification. This function reverts with confidence scores for each output class. A Softmax activation function produces a vector with four values as there are three classes of diseases and 1 class as healthy to be classified. These four values in the vector are the probability of the result. These four values sum up to one. The target class which has the highest probability is classified as the disease category. The loss function used to optimize the model is cross-entropy. The Cross-Entropy is used to calculate the distance between the output probabilities and the truth values. In order to achieve desired outcomes, the model should be as close to the actual value as possible. As part of model training, model weights are continuously adjusted with the aim of minimizing Cross-Entropy loss. Figure 9 represents the output of classification layer.

![FIGURE 9. Classification layer output.](image)

**I. GENERATION OF RICE ADVISORY**

As per the final classification result provided by the Softmax classifier, a final crop advisory is generated which provides the farmer with the support of preventive and corrective actions to be taken. The expected emulated output of the model is structured in Figure 10.

Based on the data fetched from sensors and images, the rice diseases are predicted, diagnosed, and quantified, and accordingly, the crop advisory is generated. Thus, controlling the diseases in rice grains.

**V. RESULTS AND DISCUSSION**

The training loss analysis of the Rice Transformer model is represented in Figure 11. The model starts converging at the 222nd epoch. Even though the collected dataset is comparatively small, the overfitting issue does not persist. The reason is that strategies such as augmenting the data and adding the dropout layer are used.

**A. RADAR CHART**

The Figure 12 displays the radar chart using performance measures based on the predicted four classes of rice diseases to demonstrate the superiority of the proposed framework. It can be seen that the proposed method yields the best results concerning F1-score and Accuracy parameters. The radar chart analysis concludes that the Rice Transformer model based on the concept of cross attention enhances the network’s ability to derive discriminative features for diagnosing the three classes of rice diseases and one normal class based on sensor data and image data. As a result, the proposed method can help the farmers decide on the rice crop’s illness. The proposed model outperforms all evaluation metrics in terms of classification performance, indicating that the proposed method is appealing for distinguishing the characteristics of the Healthy, blight, and blast diseases in rice crops.

**B. QUALITATIVE VISUALIZATION**

The qualitative visualization histogram is plotted in Figure 13 to demonstrate the count of the predicted classes for all the models considered in this paper. To calculate the count of the predicted categories, if the instance is classified as per its original categorization, it is counted as correctly classified. If the sample is incorrectly predicted as another category, it is estimated as an incorrectly categorized instance. It is observed that after adding a cross attention block to the proposed model, the number of classes that are predicted correctly is comparatively more significant than the models without the cross attention modules. This demonstrates that the cross attention mechanism can improve the ability of the model to focus on specific areas where infection can grow. After comparing the proposed and other model histograms, it can be seen that the cross attention module, when integrated with the proposed architecture, has brought significant gains in predicting the rice disease category, particularly for healthy, blast, and blight classes.

**C. ABLATION ANALYSIS**

Ablation analysis was carried out to investigate the effect of changing the number of encoders and decoders as well as changing the number of vector triplet dimensions represented in Table 2 and Table 3, respectively. The vector triplet dimension is set to 64, while the variation in the number of encoders and decoders was performed. The number of encoder and decoder is set to 3 while varying the vector triplet dimension. It can be observed that increasing the vector triplet dimension improves accuracy up to a certain threshold point, but beyond that point, the increase has no significant effect on accuracy. However, as the number of encoders
and decoders increases, the accuracy also increases. This demonstrates that modality fusion in conjunction with the cross attention model aids in characterizing features for better categorization.

### D. CONFUSION MATRIX

The confusion matrix for the proposed Rice Transformer model is plotted in Figure 14. The parameters used to calculate Confusion Matrix are True Negative (TN), True Positive (TP), False Positive (FP), and False Negative (FN). True negatives are the instances where the plants did not have the disease and the model also predicted that plants are not affected by the disease. True positives are cases in which the plants have diseases and the model estimated that they will have it as well. If the plants are not affected by the diseases but the model predicts that they are affected, then this is known as false positives. Similarly, if the plants do have a disease, but the model predicts that it is not affected by the disease then this is known as false negative. It can be summarized that the proposed model highly improves the performance of the prediction in terms of healthy, blight, and rice blast categories.

### E. T-DISTRIBUTED STOCHASTIC NEIGHBOR EMBEDDING (T-SNE) VISUALIZATION

The data is statistically visualized using the t-SNE method. The Figure 15 shows that the separation effect improves when a module is added to the Rice Transformer model. The t-SNE visualization of all the considered networks in the paper is performed over the testing dataset, whose size is 840 records. The visualization of the features that are learned by the last extraction layer of all the compared models is done. The index has classes 0-3, representing healthy, blight, brown spot, and blight, respectively. Figures 15(b)-(d) represent the features extracted by the early fusion model, self-attention, and rice transformer model, respectively. It is observed that the Rice Transformer approach that is based on the cross

---

**TABLE 2. Ablation analysis for change in encoder-decoder number.**

| Number of Encoder and Decoder | 1   | 2   | 3   | 4   |
|-------------------------------|-----|-----|-----|-----|
| Proposed Rice Transformer     | 93.72 | 95.53 | 95.81 | 97.38 |

**TABLE 3. Ablation analysis for change in vector triplet dimension.**

| Number of Encoder and Decoder | 32  | 64  | 128 | 256 |
|-------------------------------|-----|-----|-----|-----|
| Proposed Rice Transformer     | 92.17 | 97.38 | 96.36 | 96.05 |
attention model outperforms early fusion (based on concatenation of two modalities) and the self-attention model (based on the encoder and decoder approach). It segregates maximum features according to the rice infection classes and thus proves to be the most accurate choice for plant disease diagnosis.

**F. RECEIVER OPERATING CHARACTERISTIC CURVE**

In addition, the Receiver Operating Characteristic (ROC) curve of the proposed model along the fusion models in order to show the effectiveness of the Rice Transformer framework is represented in figure 16. It can be observed that the proposed system achieves the maximum AUC value. The closeness of the AUC score to 1 makes the model more accurate.

**G. RUNTIME PERFORMANCE**

The runtime of different algorithms in milliseconds (ms) is reported in Table 4 so that the model’s performance is compared. The runtime is recorded using a Google Colab Nvidia K80 GPU. Rice Transformer takes minimum inference time when compared to other models considered. The final value of runtime is obtained by averaging per sample inference time in milliseconds over 10 runs.

**H. COMPARISON WITH STATE-OF-THE-ART APPROACHES**

The overall accuracy of all the variants of fusion approaches deployed in this paper is calculated. The variants of fusion approaches are applied to dataset collected for the proposed work. The bar graph depicted in Figure 17 suggests that the proposed Rice Transformer works best for categorizing rice diseases compared with other fusion and self attention models.

Table 5 represents the comparison of the proposed Rice Transformer model with the approaches in the available literature. Table 5 is compared based on the attention/fusion mechanism and the architectures the authors have used in their research for various crops. Significantly less study is performed on attention mechanisms and specifically on rice crops. So latest fusion and attention work is compared

![Radar chart for Accuracy, F1-score, Recall and Precision for all predicted classes](image)
with the proposed model quoting that if attention mechanism is applied in plant disease diagnosis can achieve better results. It represents that by adding attention module a new arena is added to the research in the domain of crop disease classification as it shows improved classification results.

I. LIMITATIONS AND FUTURE WORK OF THE STUDY

Although the proposed framework outperforms the compared network models for categorizing the three rice diseases and healthy rice images, this paper has some limitations. These constraints are summarised below. Deep neural networks need substantial datasets to learn profound and discriminative features, but they lack them with the gathered dataset. There are few techniques used in this study to augment the dataset; however, a larger number can be explored. A second problem is a difficulty of collecting field-level real-time data. The studied types of rice disease species only include healthy, blast, blight, and brown spot, which restricts the applicability of the proposed algorithm to just a few rice fields. In order to overcome the limitations outlined above, a variety of agricultural sensors could be used to collect images from the field with different illuminations and capturing angles, as well as more agricultural parameters to improve future work data collection and analysis. The dataset expansion needs to be done to address the problem of small datasets by enhancing datasets in various ways, such as rotation, flipping, and cropping.

As the future prospects, agricultural research centers can collect images and sensor data to solve the problem of limited data. Additionally, extensive research can be conducted on other types of diseases. Furthermore, exploring data from multiple centers will allow the developed model to adapt robustly to the diverse environmental and soil conditions.
The sensors data can reflect the changing condition of the environment and predict the disease that crops can suffer. Farmers and agricultural experts can use scientific treatment to control the crop deterioration due to illness over time. In the future, the model can be spanned for different types of crops, and different types of diseases, making the farmer’s job easy. Also, data from many other modalities can be explored, such as soil maps, drone data, satellite images. Due to climatic and geographical challenges, future work shall collect balanced datasets for various rice diseases. The results of the proposed study show promising results for classifying infected and healthy rice leaves in real time with a variety of rice diseases. Segmenting infected leaf parts would be an exciting research topic in the future. Moreover, depending on the disease’s category and disease grade level, a fertilizer recommendation can be suggested based on the severity level of the disease.
This system allows partial automation despite using real-time data. Therefore, this is potentially another contribution for the future. The agricultural industry will benefit from these types of systems.

VI. CONCLUSION

Plant diseases are one of the most challenging problems in the agricultural domain. People around the world eat rice as a staple food. Rice Transformer, an innovative multimodal information fusion approach based on cross-attention technique, excels in classifying rice diseases. It is more accurate than individual unimodal approaches and multimodal fusion techniques. Agricultural sensor data and rice image data are inputted into the model. The rice infection categories such as Rice blast, Brown spot, bacterial blight, and healthy type is considered for the proposed system. Since these data collection samples span both modalities, images, and environmental attributes, they are unique. The feature extraction model extracts features from different modalities. These features are further passed on to the cross-attention model, where the inputs are inputted in a cross manner. Attention Score is generated out of the cross model and is further passed to pooling and softmax layers to classify rice diseases. The overall accuracy achieved is 96.9%. The Rice Transformer approach has outperformed the state-of-the-art methods that are the benchmark in rice disease classification. It proves the applicability of fusing the information from different modalities along with attention techniques to improve rice disease classification. The farmers will maintain their crop’s natural quality using such an integrated management system.

REFERENCES

[1] S. P. Mohanty, D. P. Hughes, and M. Salaléth, “Using deep learning for image-based plant disease detection,” *Frontiers Plant Sci.*, vol. 7, pp. 1–10, Sep. 2016.
[2] S. Kaur, S. Pandey, and S. Goel, “Plants disease identification and classification through leaf images: A survey,” *Arch. Comput. Methods Eng.*, vol. 26, no. 2, pp. 507–530, 2019, doi: 10.1007/S11831-018-9255-6.
[3] B. S. Bari, M. N. Islam, M. Rashid, M. J. Hasan, M. A. M. Razman, R. M. Musa, A. F. A. Nasir, and A. P. Majed, “A real-time approach of diagnosing Rice leaf disease using deep learning-based faster R-CNN framework,” *PeerJ Comput. Sci.*, vol. 7, pp. 1–27, Apr. 2021.
[4] R. R. Patil and S. Kumar, “Priority selection of agro-meteorological parameters for integrated plant diseases management through analytical hierarchy process,” *Int. J. Elect. Comput. Eng.*, vol. 12, no. 1, pp. 649–659, 2022.
[5] M. A. Zenovic, M. Karanovic, S. Sladojevic, A. Anderla, and D. Stefanovic, “Solving current limitations of deep learning based approaches for plant disease detection,” *Symmetry*, vol. 11, no. 7, p. 939, Jul. 2019.
[6] S. H. Lee, H. Goi, H. Bonnet, and A. Joly, “Attention-based recurrent neural network for plant disease classification,” *Frontiers Plant Sci.*, vol. 11, pp. 1–8, Dec. 2020.
[7] P. Moghadam, D. Ward, E. Goan, S. Jayawardena, P. Sikka, and E. Hernandez, “Plant disease detection using hyperspectral imaging,” in *Proc. Int. Conf. Digit. Image Comput.*, Techn. Appl. (DICTA), Nov. 2017, pp. 1–8.
[8] F. Görlach, E. Marks, A. K. Mahlein, K. König, P. Lottes, and C. Stachniss, “UAV-based classification of Cercospora leaf spot using RGB images,” *Drones*, vol. 5, no. 2, pp. 1–15, 2021.
[9] J. Liu and X. Wang, “Plant diseases and pests detection based on deep learning: A review,” *Plant Methods*, vol. 17, no. 1, pp. 1–18, Dec. 2021, doi: 10.1186/S13007-021-00722-9.
[10] A. Singh, S. Jones, B. Ganapathysubramanian, S. Sarkar, D. Mueller, K. Sandhu, and K. Nagasubramanian, “Challenges and opportunities in machine-augmented plant stress phenotyping,” *Trends Plant Sci.*, vol. 26, no. 1, pp. 53–69, Jan. 2021, doi: 10.1016/j.tplants.2020.07.010.
[11] S. Kumar, R. R. Patil, V. Kumawat, N. Krishnan, and S. K. Singh, “A biometric analysis of plant disease classification with artificial intelligence using convolutional neural network,” *Library Philosophy Pract.*, vol. 5777, pp. 1–14, Jul. 2021.
[12] R. U. Khan, K. Khan, W. Albattah, and A. M. Qamar, “Image-based detection of plant diseases: From classical machine learning to deep learning journey,” *Wireless Commun. Mobile Comput.*, vol. 2021, pp. 1–13, Jun. 2021.
[13] R. Patil and S. Kumar, “A biometric survey on the diagnosis of plant leaf diseases using artificial intelligence,” *Library Philosophy Pract.*, vol. 2020, pp. 1–25, Aug. 2020.
[14] G. S. Sujawat, “Application of artificial intelligence in detection of diseases in plants: A survey,” *Turkish J. Comput. Math. Educ.*, vol. 12, no. 3, pp. 3301–3305, Apr. 2021.
[15] R. R. Patil, S. Kumar, and R. Rani, “Comparison of artificial intelligence algorithms in plant disease prediction,” *Revue d’Intell. Artificielle*, vol. 36, no. 2, pp. 185–193, Apr. 2022.
R. R. Patil and S. Kumar, “Rice Transformer: A Novel Integrated Management System for Controlling Rice Diseases.” IEEE Access, vol. 10, pp. 5207–5222, 2022.

S. Zhao, Y. Peng, J. Liu, and S. Wu, “Tomato leaf disease diagnosis based on improved convolution neural network by attention module,” Agriculture, vol. 11, no. 7, p. 651, Jul. 2021.

A. Blujel, N. E. Kim, E. Arulmozi, J. K. Basak, and H. T. Kim, “A lightweight attention-based convolutional neural networks for tomato leaf disease classification,” Agriculture, vol. 12, no. 2, pp. 1–18, 2022.

W. Zeng and M. Li, “Crop leaf disease recognition based on self-attention convolutional neural network,” Comput. Electron. Agricult., vol. 172, May 2020, Art. no. 105341, doi: 10.1016/j.compag.2020.105341.

Z. Chen, R. Wu, Y. Lin, C. Li, S. Chen, Z. Yuan, S. Chen, and X. Zou, “Plant disease recognition model based on improved YOLOv5,” Agronomy, vol. 12, no. 2, p. 365, 2022. [Online]. Available: https://www.mdpi.com/2073-4395/12/2/365

G. Yang, Y. He, Y. Yang, and B. Xu, “Fine-grained image classification for crop disease based on attention mechanism,” Frontiers Plant Sci., vol. 11, pp. 1–15, Dec. 2020.

W. Ma, J. Zhao, H. Zhi, J. Shen, L. Jiao, Y. Wu, and B. Hou, “A spatial-channel collaborative attention network for enhancement of multiresolution classification,” Remote Sens., vol. 13, no. 1, pp. 1–27, 2021.

RUTUJA RAJENDRA PATIL (Member, IEEE) received the bachelor’s and master’s degrees in information technology from Pune University, Pune, India, in 2006 and 2015, respectively. She is currently pursuing the Ph.D. degree with the Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune. She is also a Research Scholar at Symbiosis International (Deemed University). Her research interests include machine learning and deep learning.

SUMIT KUMAR (Member, IEEE) received the bachelor’s degree in electronics and telecommunication from Kurukshetra University, Kurukshetra, India, in 2005, the master’s degree from the Guru Jambheshwar University of Science and Technology, Haryana, India, in 2008, and the Ph.D. degree from Jamia Millia Islamia, Delhi, India, in 2017. He is currently working as an Associate Professor at the Electronics and Telecommunication Department, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, Maharashtra, India. His research interests include antenna design, wireless networks, wireless communications, and computational intelligence.

* * *