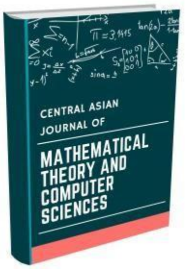

CENTRAL ASIAN JOURNAL OF MATHEMATICAL THEORY AND COMPUTER SCIENCES

https://cajmtcs.centralasianstudies.org/index.php/CAJMTCS

Volume: 06 Issue: 04 | October 2025 ISSN: 2660-5309

CENTRAL ASIAN STUDIES

.

Article

AI Voice Assistant and Caption Generation Using Convolution Neural Network and Bi LSTM

G. Rajasekaran[1], L.J. Dency Snowvin[2], K. Vaishnavi[3], D. Angelina Mary[4], S. Suman Rajest[5], M. Mohamed Sameer Ali[6]

1. Dhaanish Ahmed College of Engineering, Chennai, Tamil Nadu, India.
2. Department of Computer Science and Engineering, Dhaanish Ahmed College of Engineering, Padappai, Chennai, Tamil Nadu, India.
3. Department of Computer Science and Engineering, Dhaanish Ahmed College of Engineering, Padappai, Chennai, Tamil Nadu, India.
4. Department of Computer Science and Engineering, Dhaanish Ahmed College of Engineering, Padappai, Chennai, Tamil Nadu, India.
5. Dhaanish Ahmed College of Engineering, Chennai, Tamil Nadu, India.
6. Dhaanish Ahmed College of Engineering, Chennai, Tamil Nadu, India.

* Correspondence: rajasekaran@dhaanishcollege.in

Abstract: This research examines the physiological reactions to stress in light of the heightened stress levels observed among workers and students during the COVID-19 pandemic, particularly in distant work and virtual learning contexts. The study investigates several factors that affect stress levels in distant workers and students since it is important to understand these responses. By using ideas from psychology, physiology, and technology, the research finds the main causes of increased stress in different groups of people. The suggested approach has important effects on occupational health because it gives remote workers access to tools and information that can help them make their work environments healthier. In the same way, the system improves student well-being in virtual learning environments by giving them important support during the difficulties of remote learning. Additionally, this type of technology is useful in other areas, such as telemedicine, and it is helping to create technology-based solutions for managing stress and improving health in general. We want to help people deal with the stress of working from home and learning online by using new technologies and greater understanding of psychology. In the end, we want to help people become more resilient and healthy in the face of new obstacles.

Keywords: Technological Innovation, Educational Technology, Psychology, Physiology, Technology, Remote Work and Virtual Learning.

Citation: Rajasekaran, G., Snowvin, L. J., Vaishnavi, K., Mary, D. A., Rajest, S. S., Ali, M. M. S. AI Voice Assistant and Caption Generation Using Convolution Neural Network and Bi LSTM. Central Asian Journal of Mathematical Theory and Computer Sciences 2025, 6(4) 758-772.

Received: 10th May 2025
Revised: 16th Jun 2025
Accepted: 24th Jul 2025
Published: 07th Aug 2025

1. Introduction

In this digital age, pictures and videos are very important for talking to people, teaching them, and having fun. But for people who have trouble seeing, getting to and understanding visual information can be hard [35]. Traditional accessibility techniques have made progress in making text-based content available in other formats, but understanding visuals is still a big problem [36]. Image captioning technology has come out as a promising answer to this problem [37]. By using new deep learning techniques like convolutional neural networks (CNNs) and recurrent neural networks (RNNs), image captioning systems can automatically make descriptive captions for pictures. This makes it easier for people who are blind to access and understand visual content [38].

People who are blind or have low vision have a hard time understanding visual content, which is why new solutions like image captioning are needed [39]. Even though accessibility technologies have come a long way, many of the solutions that are already out there don't give full access to visual material [40]. Text-to-speech systems and traditional screen readers are great at turning text into audio, but they have a hard time understanding and describing what visuals mean [41]. As a result, visually impaired individuals are often precluded from fully engaging with visual media, instructional materials, and online content. To solve this problem, we need to create strong image captioning systems that can correctly assess visual content and make captions that fit the user's demands and preferences [42].

Each deep learning model requires a set of data to roduce reliable predictions. Flickr8k, Flickr30k, and Microsoft COCO are three datasets that are often used to test and compare image captioning algorithms [43]. The model uses the Flickr8K dataset, which has more than 8000 photos that have been labeled by Amazon Mechanical Turk and five separate statements [44]. The goal of this initiative is to make the digital world more open and welcoming to everyone [45]. People who are blind or have low vision should have the same access to information and media, including visual content [46]. We want to give blind people the same access to visual media as sighted people by creating an automatic picture captioning system [47]. Adding informative captions to photographs not only makes them easier to reach, but it also makes people feel like they belong and are equal in the digital world [48]. Also, new developments in deep learning give us new chances to make picture captioning systems that are more advanced and accurate, which makes us even more determined to do this [49]. In the end, the goal is to use technology to improve the lives of those who are blind or have low vision and to make society more welcoming so that everyone may fully participate in the digital world [50].

Literature Survey

There has been a lot of interest in combining language and vision in recent years, mostly because of the amazing progress made in artificial intelligence (AI) and related subfields, such as machine learning, computer vision, and natural language processing (NLP) [1]. A lot of this growth is due to the rise of deep learning methods, especially in artificial neural networks, which have made it much easier to combine language and vision tasks. In this survey, we examine ten significant tasks that integrate language and vision domains. We examine the issue formulation, approaches utilized, available datasets, and evaluation measures for each task, comparing the outcomes with state-of-the-art procedures [2]. Our poll is different from previous ones since it looks at all sorts of visual content (such images or videos) and all types of tasks. We want to give a full picture of the area where language and vision meet, covering a wide range of tasks and modes [4]. We also look beyond the field's current condition to see what might happen in the future and what new areas of research might be worth exploring. By combining ideas from different research papers, we hope to spark new ways of thinking and encourage people from other fields to work together [5]. We hope this poll will help solve current problems, spark the creation of new applications, and support a lively research community in the interesting field of language and visual integration. We hope that this paper will help AI continue to grow and change society in important ways [6].

The research examines finite-time convergent complex-valued zeroing neural networks (IFTCVZNN), providing innovative perspectives on the real-time resolution of the time-varying reciprocal of complex matrices [7]. The study examines two enhanced IFTCVZNN models, utilizing two separate processing methodologies for intricate computations related to nonlinear activation functions. The study also looks into a new nonlinear activation function to improve the overall performance of the two IFTCVZNN models [8]. When compared to other complex-valued neural networks that converge within a certain amount of time, the suggested IFTCVZNN models, which use a new activation function, show better finite-time convergence and less strict upper bounds [9]. Theoretical evaluations, supported by numerical simulations, confirm that the maximum convergence time, as determined by Lyapunov stability, roughly corresponds to the actual convergence time [10]. This literature enhances the comprehension of complex-valued neural networks, especially in finite-time convergence contexts, and presents encouraging prospects for real-time applications in intricate matrix computations [3].

The automatic development of English descriptions for remote-sensing photographs has attracted considerable interest in the field of remote sensing [11]. Recent deep learning methods for attention-based captioning can make words while highlighting the locations of objects in the image [12]. Nevertheless, traditional attention-based methods often depend on coarse-grained and unstructured attention units, which means they don't fully take use of the organized spatial relationships that are naturally present in the semantic content of remote sensing images [13]. Many methods for annotating remote sensing photos borrow significantly from computer vision techniques, even though remote sensing images have their own unique structural features that set them apart from natural images. These methods often ignore knowledge that is relevant to the field [14]. To fill this gap, a new fine-grained, structured attention-based method has been suggested to make use of the natural structural features of semantic content in high-resolution remote sensing images [16]. This method seeks to yield more precise descriptions while creating pixel-wise segmentation masks for semantic content. The suggested solution allows for the simultaneous training of segmentation and captioning tasks within a cohesive framework, eliminating the need for pixel-level annotations. The suggested method has been tested on three benchmark datasets for remote sensing image captioning, including thorough ablation studies and parameter assessments [17]. A comparison with the best methods shows that our method has better captioning accuracy and can make both high-resolution and semantically relevant segmentation masks at the same time. This study, published in the IEEE Transactions on Geoscience and Remote Sensing (Volume: 60), signifies a notable progression in the comprehension of remote sensing imagery, providing a comprehensive methodology for producing descriptive language in conjunction with pixel-wise segmentation masks, thus enhancing the analysis and interpretation of remote sensing data [18].

In the upcoming Industry 5.0, people will be very important to the human-cyber-physical system. Because incorrect work material causes stress and has negative impacts on human performance, engineering applications look for a physiological indicator that can be used to keep an eye on workers' health while they are on the job [19]. This way, the work content may be changed as needed. The principal objective of this study is to evaluate the validity and reliability of heart rate variability (HRV) as an indication of acute work-content-related stress (AWCRS) in real-time during industrial labor [20]. Second, we seek to present an expanded perspective on the application of HRV as a stress indicator within this setting. Ways: A search was performed in Scopus, IEEE Xplore, PubMed, and Web of Science from January 1, 2000, to June 1, 2022. Eligible articles are examined for study design, population, evaluation of AWCRS, and its correlation with HRV. Results: Fourteen studies satisfied the inclusion criteria. There was no randomized control experiment (RCT) to find out if there is a link between AWCRS and HRV. There were five observational studies done [21]. AWCRS and HRV were assessed in nine additional investigations; however, their correlations were not examined. The findings indicate that HRV does not completely represent the AWCRS during labor, and measuring the impact of AWCRS on HRV in a genuine manufacturing setting presents challenges. The research is inadequate to draw a definitive conclusion on the HRV diagnostic function as a measure of human worker status [15].

People's thoughts and feelings are shown in their facial expressions. It gives the viewer a lot of social clues, like what to pay attention to, what to want, what to feel, and what to do [23]. People think of it as a powerful way to communicate without words. Examining these expressions provides a far deeper understanding of human behavior. AI-based Facial Expression detection (FER) has emerged as a significant study domain in recent years, with applications in dynamic analysis, pattern detection, interpersonal interaction, mental health monitoring, and numerous others [24]. But because of the COVID-19 epidemic, which has led to some global push toward online platforms, there has been a strong need to come up with a new FER analysis framework to deal with the growing amount of visual data from videos and photos [25]. Additionally, the emotional facial expressions of children, adults, and elderly individuals differ, which must also be taken into account in FER study. But it doesn't give a full picture of the literature that shows prior effort and suggests future directions [26]. The authors of this study have thoroughly assessed AI-based FER approaches, encompassing datasets, feature extraction techniques, algorithms, and recent advancements in their applications for facial expression recognition [27]. This is the only review paper that covers all elements of FER for different age groups, and it will have a big effect on the research community in the next few years [22].

Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks are two types of networks that are often employed in the current system for picture captioning. We use CNNs to get features from images, and then we use LSTMs to make sequential captions from those features [28]. CNNs are great at getting visual information from pictures and capturing key spatial information such shapes, textures, and objects. These attributes give the model a lot of information about the image content, which helps it understand the visual context. LSTM networks, on the other hand, are very important for making sequential captions since they process the image features that have been taken out [29]. LSTMs are great at finding long-range dependencies in sequential data and can make captions for photos that make sense and fit the context. The current method effectively generates descriptive and contextually relevant captions for different images by combining CNNs for feature extraction and LSTMs for sequential caption production [30].

The suggested approach uses advanced models like VGG18 and Bidirectional Long Short-Term Memory (BiLSTM) networks to add captions to images. VGG18 is chosen because it can efficiently collect both low-level and high-level information [32]. It is noted for being good at extracting rich visual aspects from photos. Adding VGG18 to the proposed system makes it easier to interpret complicated visual contexts and improves the quality of feature extraction. BiLSTMs are also chosen because they can get more complete contextual information by looking at data from both the past and the future [33]. This bidirectional processing helps the model make the captions it generates more coherent and fluent, and it also makes it better at handling long-range dependencies in sequential data. The suggested approach seeks to improve the accuracy and contextual relevance of image captions by utilizing the strengths of VGG18 and BiLSTMs [34]. This enhancement ultimately facilitates improved accessibility and comprehension of visual content for those with visual impairments, thereby enabling them to engage more comprehensively with the visual realm [31].

2. Materials and Methods

The first stage in the process is to get the dataset ready to train the model that will write captions for images. Most of the time, the dataset is made up of pairs of photographs and their captions. We preprocess each image-caption combination to get features and change the text into a format that the model can learn from. A pre-trained Convolutional Neural Network (CNN) model like VGG16 or VGG19 is used to get visual features from photos [52]. These CNN models have been trained on big datasets and can pick up on high-level features in images very well. The CNN model's output is a feature vector that shows what is in the image. Text preparation methods are used on captions to clean up the text and break it up into tokens [53]. This means taking out special characters, making the text all lowercase, and breaking it up into separate words or tokens. Also, padding is used to make sure that all of the captions have the same length [54].

After the data has been cleaned up, the next step is to train the picture captioning model. The model design usually has two primary parts: an encoder for images and a decoder for text [55]. The image encoder takes care of processing the picture features that were taken out of the data during preprocessing [56]. The suggested system uses VGG18 as the image encoder to get visual attributes from the images. After that, these features are sent to the text decoder part. The text decoder uses Bidirectional Long Short-Term Memory (BiLSTM) networks to make captions depending on the attributes of the image [57]. BiLSTMs are good at making captions that make sense and fit the context because they can remember long-range dependencies in sequential data [58]. The model learns how to make captions by minimizing a loss function that looks at the difference between the predicted and ground truth captions in the training set. Backpropagation and gradient descent optimization are two ways to do this [59].

After the Model has been trained, it may make captions for fresh pictures. To get visual features from the image, you have to pass it through the trained model and use the image encoder (VGG18). The text decoder (BiLSTM) then uses these features to make a string of words that make up the caption [60]. The process of making captions is iterative, meaning that each word is predicated on the words that came before it. The model keeps making words until it predicts an end-of-sequence token or reaches the maximum length for a caption [61]. The image captioning paper uses advanced methods like CNNs (VGG18) for extracting picture features and BiLSTMs for creating sequential captions. These methods include data preparation, model training, and caption synthesis [62]. These methods let the model make captions for different images that are descriptive and relevant to the context. This makes visual content easier to access and understand for people who are blind or have low vision [51].

Adding Pyttsx3 to image captioning systems is a big step forward in making things easier for people who can't see. Pyttsx3 lets users understand visual content by turning written captions into speech that sounds natural. This closes the gap between how people perceive visual and aural information [63]. Pyttsx3 is a flexible Python toolkit for converting text to voice (TTS). It provides a strong framework for making speech from text inputs with great clarity and expressiveness [64]. First, adding Pyttsx3 to the image captioning pipeline requires a number of steps to make sure it works well and fits in with the rest of the system. The first step for developers is to install Pyttsx3 in the Python environment. Then, they set up options like voice selection, speech rate, and loudness modulation based on what the user wants and what is best for accessibility [65]. Pyttsx3 is easy to use for converting text to speech after it is installed and set up. This makes it easy to turn text into speech. One of the best things about Pyttsx3 is that it can be customized, which lets developers change the speech synthesis settings to fit their own needs [66]. Users can choose from a number of different voices, change the speed and loudness of speech, and add pauses or emphasis to make their speech more expressive [67].

This flexibility makes sure that the voice output from Pyttsx3 meets the needs and preferences of those who are blind or have low vision, which improves the overall user experience [68]. Also, Pyttsx3 has strong error-handling features that make sure speech synthesis keeps going even when there are problems or interruptions. Developers can add error-handling methods to handle exceptions smoothly and keep speech output going, which makes the TTS conversion process more reliable and robust [69]. This strength is very important for making sure that users have a smooth experience, especially those who depend on speech output to get information [70].

Adding Pyttsx3 to the image captioning system's user interface makes it even easier to use by making it easy to switch from text captions to speech output [71]. After creating appropriate captions for pictures, they are sent to the user through audio output. This integration makes the user experience smoother, allowing those who are blind to fully interact with visual content without having to rely just on text descriptions [72]. Pyttsx3 has a lot of potential to improve accessibility in many different areas and applications, beyond just image captioning systems. Pyttsx3 can be used to make information more accessible and welcoming for people with different requirements and preferences. For example, it can be used to create assistive technologies for people with impairments, interactive educational tools, and virtual assistants [73]. As TTS technology keeps getting better, Pyttsx3 stays at the cutting edge of innovation, pushing for accessibility and inclusion in the digital age [74]. In conclusion, adding Pyttsx3 to image captioning systems is a big step forward in making things easier for people who are blind or have low vision. As technology keeps changing, Pyttsx3 is ready to play a key role in pushing for accessibility and inclusivity, giving people with disabilities the tools they need to easily and confidently use and interact with digital material [75].

Design Process

User Interface (UI): The web app will have a simple and easy-to-use interface that lets users upload pictures and get meaningful captions [76]. There will be parts of the UI for uploading pictures and showing captions. The Python web framework Flask is used for this. Backend Server: The backend server will be in charge of the main parts of the image captioning system, such as processing images and making captions. It will take image uploads from the UI, process them with the image processing module (using VGG18 for feature extraction), and then use the caption generation module (using BiLSTM networks) to make captions.

Data Storage: The web app might have a database or file storage system to keep track of uploaded photographs, generated captions, and user preferences. This part of data storage will make sure that user interactions are permanent and can be found later (Figure 1).

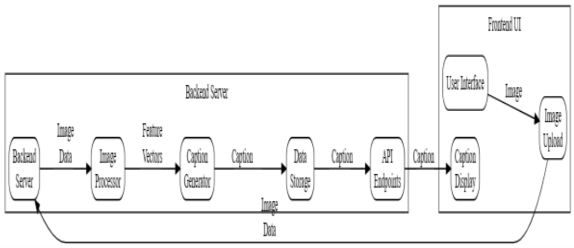

Figure 1. Data Flow Diagram

This part is what the web app's user interface looks like. Users use this interface to upload pictures and see captions. The "User Interface" node shows the graphical interface that users utilize to interact with the program [77]. The "Image Upload" node shows how users can upload pictures to the web app [78]. The "Caption Display" node shows where the processed captions are shown to the users. This part stands for the server that handles image data and makes captions [79]. The "Backend Server" node includes the functionality on the server side that handles requests and manages the processing of data.

Architecture Diagram

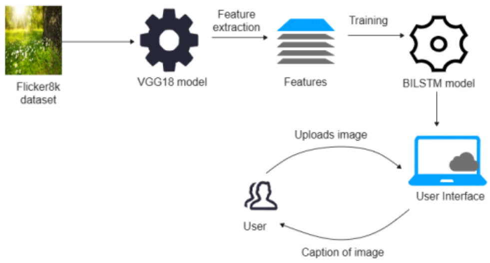

Figure 2. Architecture Diagram

Figure 2 shows a web application for adding captions to images. It has a frontend UI and a backend server. The front-end UI lets users upload photographs and see captions.

Images that are uploaded are transferred to the backend server over API endpoints [80]. There, they are processed using VGG18 for feature extraction and BiLSTM networks for caption generation. The data storage component keeps the captions so they can be used again later [81]. The backend server has API endpoints that let the frontend UI talk to it, making it easy for data to move between the two parts [82]. Overall, the architecture makes it easier to analyze images, make captions, and store them, which improves the user experience of the image captioning software.

Sequence Diagram

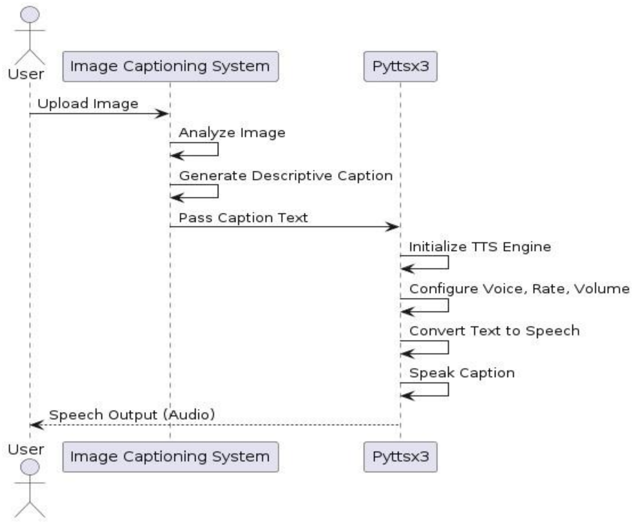

Figure 3. Sequence Diagram

User Uploads Image: The first step in the interaction is for the user to upload an image to the image captioning system. When you do this, the system looks at the uploaded picture and makes a caption that describes it [83]. Image Analysis and Caption Generation: The image captioning system looks at the picture you submitted, finds important parts of it, and then makes a descriptive caption based on what it sees. After the caption is made, it is sent to Pyttsx3 to be turned into speech (Figure 3).

Text-to-Speech Conversion: Pyttsx3 receives the generated caption text from the image captioning system and proceeds with the text-to-speech conversion procedure [84]. It sets up the voice, pace, and volume settings for the text-to-speech engine and changes the text caption into speech output [85].

System Implementation

Anaconda Navigator is a sophisticated and easy-to-use platform that makes it easier to manage and deploy data science and machine learning workflows. This program has a clean and easy-to-use interface that makes it easy for experts and hobbyists to work with the complexities of data analysis, machine learning development, and scientific computing. Anaconda Navigator is the way to get to the huge Anaconda distribution [86]. This version of Python comes with free and open-source tools and packages that are necessary for machine learning and data science work. The main dashboard of Anaconda Navigator gives users a simple, all-in-one place to find and use different tools and environments. It gives a clear picture of the apps that are available, like Jupyter Notebooks, JupyterLab, and Spyder. Jupyter Notebooks is a popular interactive computing environment that lets people make and share live code, equations, visualizations, and narrative text documents. JupyterLab is an extendable platform that lets you explore and analyze data interactively. Spyder, on the other hand, is an IDE made specifically for scientific computing and data analysis in Python.

One of the best things about Anaconda Navigator is how easy it is to manage environments. Users can make separate environments with their own dependencies, packages, and versions of Python. This makes it easier to work on and test papers with various needs without worrying about dependencies getting in the way. Anaconda Navigator's ability to manage environments makes it easier to exchange papers between systems and makes them more likely to work the same way every time. Anaconda Navigator also makes it easier to install and update Python packages and dependencies. Users can browse, install, and update packages from the large Anaconda repository or other places. This makes it easier to build up the complicated libraries and tools needed for advanced machine learning and data analysis jobs. Anaconda Navigator is also very important for helping data scientists work together and share what they know. The platform has a lot of ways to learn, like tutorials, documentation, and community forums. This encourages people to keep learning and makes it easier for them to share ideas and best practices. Data scientists, researchers, and developers that want a single, efficient platform for their Python-based data science and machine learning operations will find Anaconda Navigator to be a useful and necessary tool. Its straightforward interface, environment management features, and access to a broad selection of packages make it a great asset for beginners and experienced professionals navigating the intricacies of modern data analysis and scientific computing challenges. Anaconda Navigator is a strong and easy-to-use tool for the wide and changing field of data science. It can be used to look at datasets, build machine learning models, or work with other people.

Anaconda is a free and open-source version of the Python and R programming languages for scientific computing (data science, machine learning, processing massive amounts of data, predictive analytics, etc.). Its goal is to make it easier to manage and install packages. The package management system "Conda" takes care of package versions. The Anaconda distribution has more than 12 million users and comes with more than 1400 popular data science packages that work on Windows, Linux, and MacOS. Anaconda distribution comes with more than 1,400 packages, as well as the Conda package and virtual environment manager called Anaconda Navigator. This means you don't have to learn how to install each library separately. You may install the open-source package scan from the Anaconda repository using the command or with the pip install command that comes with Anaconda

Technologies Used

Tensor Flow: Google made and released Tensor Flow, a Python library for fast numerical computing. It is a base library that lets you build Deep Learning models directly or with wrapper libraries that make the process easier on top of TensorFlow.

A num-py array is a grid of items of the same type, each of which is indexed by a tuple of non-negative integers. The rank of the array is the number of dimensions it has. A tuple of integers that tells you the size of the array along each dimension is what gives it its shape.

SciPy (Scientific Python) is often mentioned in the same breath as Numpy

SciPy adds more useful functions to Numpy for things like minimization, regression, Fourier transformation, and a lot more. There are two older Python modules that deal with arrays that Numpy is built on. One of these is Numeric. Numeric is a Python module for high-performance numeric calculation, just like Numpy. However, it is no longer useful. Numarray is another predecessor of Numpy. It is a complete rewrite of Numeric, however it is likewise no longer supported. Numpy is a combination of the two; it is based on the code of Numeric and the properties of num array. Image captioning is an interesting mix of computer vision and natural language processing that aims to automatically provide text descriptions for pictures. Old-fashioned methods for image captioning used hand-made features and rule-based systems, which typically had trouble capturing the complexity and variety of natural language descriptions. But deep learning, especially Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), has made a big difference in the discipline. The goal of this paper is to build an image captioning system that uses deep learning techniques. We will train the model using a Bidirectional LSTM (BiLSTM) architecture and use the pre-trained CNN model

VGG16 to get picture features. Flask, a lightweight web framework in Python, will be used to build the app. This will let users upload photographs and get captions from the model in real time.

3. Results and Discussion

The Flickr8K dataset is a well-known benchmark dataset for picture captioning tasks that was chosen for this study. There are 8,000 pictures in it, and each one has five human-written captions that give a lot of detail about different scenes and things. The dataset's richness and variety make it a good choice for training a strong image captioning model that can produce meaningful captions for a wide range of images (Figure 4).

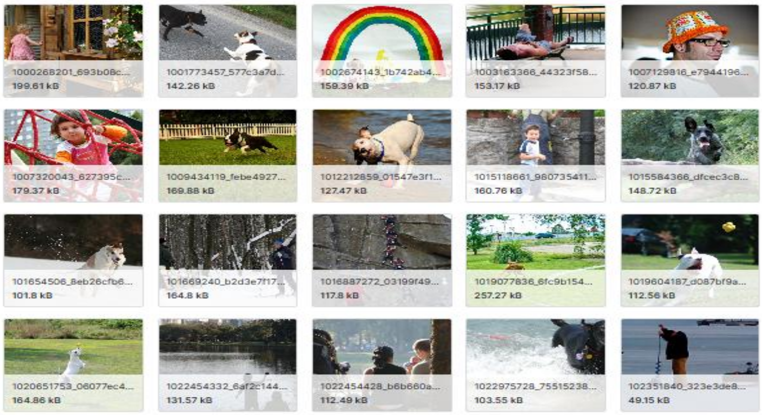

Figure 4. Flicker8K dataset

```
image,caption
1000268201_693b08cb0e.jpg,A child in a pink dress is climbing up a set of stairs in an entry way .
1000268201_693b08cb0e.jpg,A girl going into a wooden building .
1000268201_693b08cb0e.jpg,A little girl climbing into a wooden playhouse .
1000268201_693b08cb0e.jpg,A little girl climbing the stairs to her playhouse .
1000268201_693b08cb0e.jpg,A little girl in a pink dress going into a wooden cabin .
1001773457_577c3a7d70.jpg,A black dog and a spotted dog are fighting
1001773457_577c3a7d70.jpg,A black dog and a tri-colored dog playing with each other on the road .
1001773457_577c3a7d70.jpg,A black dog and a white dog with brown spots are staring at each other in the stree
1001773457_577c3a7d70.jpg,Two dogs of different breeds looking at each other on the road .
1001773457_577c3a7d70.jpg,Two dogs on pavement moving toward each other .
1002674143_1b742ab4b8.jpg,A little girl covered in paint sits in front of a painted rainbow with her hands ir
1002674143_1b742ab4b8.jpg,A little girl is sitting in front of a large painted rainbow .
1002674143_1b742ab4b8.jpg,A small girl in the grass plays with fingerpaints in front of a white canvas with a
1002674143_1b742ab4b8.jpg,There is a girl with pigtails sitting in front of a rainbow painting .
1002674143_1b742ab4b8.jpg,Young girl with pigtails painting outside in the grass .
1003163366_44323f5815.jpg,A man lays on a bench while his dog sits by him .
1003163366_44323f5815.jpg,A man lays on the bench to which a white dog is also tied .
1003163366_44323f5815.jpg,a man sleeping on a bench outside with a white and black dog sitting next to him .
1003163366_44323f5815.jpg,A shirtless man lies on a park bench with his dog .
1003163366_44323f5815.jpg,man laying on bench holding leash of dog sitting on ground
1007129816_e794419615.jpg,A man in an orange hat starring at something .
1007129816_e794419615.jpg,A man wears an orange hat and glasses .
1007129816_e794419615.jpg,A man with gauges and glasses is wearing a Blitz hat .
1007129816_e794419615.jpg,A man with glasses is wearing a beer can crocheted hat .
1007129816_e794419615.jpg,The man with pierced ears is wearing glasses and an orange hat .
1007320043_627395c3d8.jpg,A child playing on a rope net .
1007320043_627395c3d8.jpg,A little girl climbing on red roping .
1007320043_627395c3d8.jpg,A little girl in pink climbs a rope bridge at the park .
1007320043_627395c3d8.jpg,A small child grips onto the red ropes at the playground .
1007320043_627395c3d8.jpg,The small child climbs on a red ropes on a playground .
1009434119_febe49276a.jpg,A black and white dog is running in a grassy garden surrounded by a white fence .
```

Figure 5. Caption

Data Preprocessing

Image Feature Extraction: We use the VGG16 CNN architecture to preprocess the images and get high-level features before training the image captioning model. VGG16 is a deep convolutional neural network that is well-known for how well it works at classifying images. We can get a compact representation of each image that shows its visual features by taking out the fully connected layers and using the output of the last convolutional layer.

Preprocessing the Captions: The captions in the Flickr8K dataset are preprocessed so that they can be used to train the model. This involves tokenization, which breaks each caption down into its own words or tokens, and making vocabulary mappings. Also, specific tokens like <start> and <end> are added to show where captions start and end.

Dataset Split: The Flickr8K dataset is divided into training and validation sets to make it easier to train and test models. A conventional split ratio, such 80% for training and 20% for validation, is usually utilized. The training set is used to change the model's parameters, while the validation set is used to keep an eye on how well the model is doing and stop it from overfitting.

Training Procedure: The model is trained to minimize a loss function, like cross-entropy loss, between the predicted captions and the real captions. Backpropagation and gradient descent-based optimization methods, such Adam Optimizer, are used to iteratively update the model parameters during optimization. To make training more stable, techniques like instructor forcing, where the model has to apply the ground truth preceding word during training, might be used.

Optimization: To stop overfitting and speed up convergence, many optimization methods are used. Learning rate scheduling changes the learning rate during training to change how big the steps are when updating parameters. Early stopping ends training when the model's performance on the validation set stops getting better. This keeps the model from overfitting to the training data. Gradient clipping also stops gradients from getting too big, which stops exploding gradients during training (Figure 5).

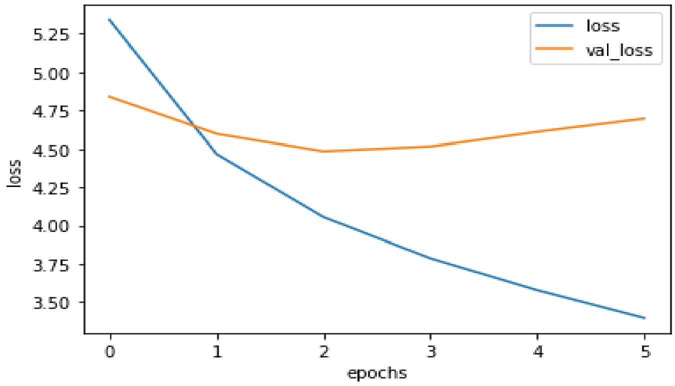

Figure 6. Result

Flask Framework: The picture captioning app was made with Flask, a lightweight and flexible web framework for Python. Flask has the tools and packages we need to build web apps with as little extra code as possible, which makes it the best choice for our paper. We can rapidly make a web-based interface for working with the image captioning model by using Flask**.**

User UI: The app's UI is meant to be easy to understand and utilize. Users can easily upload pictures and get descriptions that the model makes up on the spot. The interface might have things like drag-and-drop picture upload, progress bars, and responsive design so that it works on all devices.

Backend Integration: The Flask app works perfectly with the picture captioning model on the backend. When the app gets a request to upload an image, it first scans the image, then uses VGG16 to find features, and finally uses the trained BiLSTM model to make captions. The web interface then shows the user the created captions, which completes the entire captioning process.

4. Conclusion

The incorporation of Bidirectional Long Short-Term Memory (BiLSTM) networks into the image captioning system signifies a substantial progress in improving accessibility for those with visual impairments. The system has shown that it can use cutting-edge deep learning technology to analyze images in great detail, pick out important aspects, and create descriptive captions in simple English. The model combines deep learning and artificial intelligence to precisely grasp the information of images and put it into a format that everyone can understand, making it easier for people with visual impairments to understand. The system has been thoroughly tested and evaluated throughout the development process, confirming that it works, is reliable, and is easy to use, making it ready for use in the real world.

Future Enhancements

The present version of the picture captioning system that uses BiLSTM networks has made a lot of progress in making it easier to use, but there are still many ways to make it even better and more useful. Some possible areas for future growth are better caption quality, which means that algorithms for generating captions are always being improved to make them more accurate, coherent, and relevant. Using advanced natural language processing (NLP) techniques and semantic understanding to make descriptions that are more complex and relevant to the situation. Multimodal Fusion: Investigating methods for incorporating supplementary modalities, including audio, text, and contextual information, to enhance the comprehension and depiction of images. Attention mechanisms and fusion methods are two types of multimodal fusion that can help the system capture different parts of an image's content. Interactive User Interfaces: Making user interfaces that are easy to use and interactive, especially for people who are blind or have low vision. Adding ways for users to give feedback, use gestures, and control commands with their voices to make things easier to use and get to. Personalization and Adaptation: Setting up ways to tailor the user experience to each person's needs, interests, and accessibility demands. Adaptive learning algorithms can change how the system works and what it outputs in real time to meet the needs of each user. Real-Time Captioning: Looking into ways to make it possible to add captions to live video streams or camera feeds in real time so that those who are blind can get immediate descriptions of what is going on around them. Ethical Considerations: Dealing with ethical issues like privacy, bias, and justice in systems that add captions to images. Put in place protections to reduce the chance of bias in caption production and make sure that AI technologies are used responsibly. Working with visually impaired communities and advocacy groups to get comments, insights, and examples of how the product is used in real life. Engaging users in co-design processes and participatory design workshops to make sure the system satisfies their needs and fits with what they want.

REFERENCES

[1] G. Kulkarni, V. Premraj, S. Dhar, et al., “Baby talk: Understanding and generating simple image descriptions,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition, IEEE Computer Society, United States of America, 2011.

[2] A. F. Biten, L. Gomez, and D. Karatzas, “Let there be a clock on the beach: Reducing object hallucination in image captioning,” in Proc. IEEE/CVF Winter Conf. Applications of Computer Vision, Florida, United States of America, 2022.

[3] L. A. Hendricks, K. Burns, K. Saenko, T. Darrell, and A. Rohrbach, “Women also snowboard: Overcoming bias in captioning models,” in Proc. European Conf. Computer Vision (ECCV). Munich, Germany, 2018.

[4] T. Yao, Y. Pan, Y. Li, and T. Mei, “Exploring visual relationship for image captioning,” in Proc. European Conf. Computer Vision (ECCV), Munich, Germany, 2018.

[5] S. Herdade, A. Kappeler, K. Boakye, and J. Soares, “Image captioning: Transforming objects into words,” in Advances in Neural Information Processing Systems, Vancouver, Canada, 2019.

[6] S. A. Karthik, S. B. Naga, G. Satish, N. Shobha, H. K. Bhargav, and B. M. Chandrakala, “AI and IoT-infused urban connectivity for smart cities,” in Future of Digital Technology and AI in Social Sectors, D. Ertuğrul and A. Elçi, Eds. IGI Global Scientific Publishing, 2025, pp. 367–394.

[7] S. Rashmi, B. M. Chandrakala, D. M. Ramani, and M. S. Harsur, “CNN based multi-view classification and ROI segmentation: A survey,” Global Transitions Proceedings, vol. 3, no. 1, pp. 86–90, 2022.

[8] K. S. N. S. Nischal, N. S. Guvvala, C. Mathew, G. C. S. Gowda, and B. M. Chandrakala, “A survey on recognition of handwritten ZIP codes in a postal sorting system,” International Research Journal of Engineering and Technology (IRJET), vol. 7, no. 3, pp. 1–4, May 2020.

[9] B. M. Chandrakala and S. C. Linga Reddy, “Proxy re-encryption using MLBC (Modified Lattice Based Cryptography),” in Proc. Int. Conf. Recent Advances in Energy-efficient Computing and Communication (ICRAECC), Nagercoil, India, 2019, pp. 1–5.

[10] H. S. Supriya and B. M. Chandrakala, “An efficient multi-layer hybrid neural network and optimized parameter enhancing approach for traffic prediction in Big Data Domain,” The Journal of Special Education, vol. 1, no. 43, pp. 94–96, 2022.

[11] R. Sushmitha, A. K. Gupta, and B. M. Chandrakala, “Automated segmentation technique for detection of myocardial contours in cardiac MRI,” in Proc. Int. Conf. Communication and Electronics Systems (ICCES), Coimbatore, India, 2019, pp. 986–991.

[12] K. Shanthala, B. M. Chandrakala, N. Shobha, and D. D., “Automated diagnosis of brain tumor classification and segmentation of MRI images,” in Proc. Int. Conf. Confluence of Advancements in Robotics, Vision and Interdisciplinary Technology Management (IC-RVITM), Bangalore, India, 2023, pp. 1–7.

[13] B. M. Chandrakala et al., “Harnessing online activism and diversity tech in HR through cloud computing,” in Future of Digital Technology and AI in Social Sectors, D. Ç. Ertuğrul and A. Elçi, Eds. IGI Global Scientific Publishing, 2025, pp. 151–182.

[14] A. Navya and B. M. Chandrakala, “The effective dashboard to control the intrusion in the private protection of the cloudlet based on the medical mutual data using ECC,” in Proc. Int. Conf. Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2018, pp. 538–543.

[15] B. M. Chandrakala and S. C. Lingareddy, “Secure and efficient bi-directional proxy re-encryption technique,” in Proc. Int. Conf. Control, Instrumentation, Communication and Computational Technologies (ICCICCT), Kumaracoil, India, 2016, pp. 88–92.

[16] V. Hiremath, “Quantum Networking: Strategic Imperatives for Enterprises and Service Providers in the Emerging Quantum Era,” Journal of Computational Analysis and Applications (JoCAAA), vol. 31, no. 3, pp. 617–631, Dec. 2023.

[17] V. Hiremath, “Real-Time BGP Monitoring with BMP and Streaming Telemetry,” International Journal of Environmental Science, vol. 11, no. 1s, pp. 1109–1115, Mar. 2025.

[18] N. J. Maiti, S. Ganguly, K. Choowongkomon, S. Seetaha, S. Saehlee, and T. Aiebchun, "Synthesis, in vitro Anti-HIV-1RT evaluation, molecular modeling, DFT and acute oral toxicity studies of some benzotriazole derivatives," J. Struct. Biol., vol. 216, no. 2, p. 108094, 2024.

[19] N. J. Maiti and S. Ganguly, "Synthesis, spectral analysis, antimicrobial evaluation, molecular modelling, DFT, TD-DFT and SAR studies of novel 4,5,6,7-tetrabromo-1H-benzo[d][1,2,3]triazole derivatives," ChemistrySelect, vol. 9, no. 36, p. e202401746, 2024.

[20] D. K. Arora et al., “An in vitro assessment of microleakage of pit and fissure sealants and restorative materials using dye penetration method,” Journal of Pharmacy and Bioallied Sciences, Feb. 2025.

[21] R. Nagar et al., “In vitro analysis of compressive strength of three different aesthetic restorative materials,” Journal of Pharmacy and Bioallied Sciences, Feb. 2025.

[22] N. Maiti et al., “Assessment of the efficacy of photobiomodulation (PBM) therapy in periodontal treatment: a longitudinal study,” Journal of Pharmacy and Bioallied Sciences, vol. 16, no. Suppl 3, pp. S2449–S2451, Jul. 2024.

[23] N. J. Maiti and S. Ganguly, "Some new benzotriazole derivatives: Synthesis, antimycobacterial evaluation, antimicrobial efficacy, ADME studies, and molecular docking studies," Indian Journal of Heterocyclic Chemistry, vol. 33, no. 3, pp. 385–392, 2023.

[24] N. J. Maiti, S. Ganguly, B. Sarkar, and R. Saha, "New benzotriazole derivatives: Synthesis, biological assessment, in vivo oral toxicity analysis, docking studies, molecular dynamics, and ADME profiling," Indian Journal of Heterocyclic Chemistry, vol. 33, no. 4, pp. 489–497, 2023.

[25] N. J. Maiti, "A comprehensive review on analytical techniques for the quantification of pharmaceutical compounds in biological matrices," Journal of Cardiovascular Research, vol. 15, no. 9, 2024.

[26] A. Vahora, R. Patel, B. Goradiya, and A. Desai, ‘Heart beat monitoring and wireless data logging using arm cortex A8’, International Journal on Recent and Innovation Trends in Computing and Communication, vol. 2, no. 8, pp. 2321–2325, 2014.

[27] A. Vahora, B. Goradiya, D. Parikh, and A. Shah, ‘Designing a Model for Traffic Rule Violation at Railway Track Using Raspberry Pi in Indian Context’, International Journal of Latest Technology in Engineering,Management & Applied Science, vol. 6, no. 6, pp. 122–125, 2017.

[28] A. Vahora and K. Pandya, 'Implementation of cylindrical dielectric resonator antenna array for Wi-Fi/wireless LAN/satellite applications', Progress in Electromagnetics Research M, vol. 90, pp. 157–166, 2020.

[29] A. Vahora and K. Pandya, 'Triple Band Dielectric Resonator Antenna Array Using Power Divider Network Technique for GPS Navigation/Bluetooth/Satellite Applications', International Journal of Microwave and Optical Technology, vol. 15, no. 4, pp. 369–378, 2020.

[30] A. Vahora and K. Pandya, 'A miniaturized cylindrical dielectric resonator antenna array development for GPS/Wi-Fi/wireless LAN applications', e-Prime-Advances in Electrical Engineering, Electronics and Energy, vol. 2, p. 100044, 2022.

[31] D. Sumathi and P. Poongodi, "Scheduling Based on Hybrid Particle Swarm Optimization with Cuckoo Search Algorithm in Cloud Environment," IIOAB Journal, vol. 7, no. 9, pp. 358-366, 2016.

[32] D. Sumathi and P. Poongodi, "Secure medical information processing in cloud: Trust with swarm based scheduling," Journal of Medical Imaging and Health Informatics, vol. 6, no. 7, pp. 1636-1640, 2016.

[33] D. Sumathi and P. Poongodi, "An improved scheduling strategy in cloud using trust based mechanism," Int. J. Comput. Electr. Autom. Control Inf. Eng, vol. 9, no. 2, pp. 637-641, 2015.

[34] V. B. Gowda, M. T. Gopalakrishna, J. Megha, and S. Mohankumar, "Foreground segmentation network using transposed convolutional neural networks and up sampling for multiscale feature encoding," Neural Netw., vol. 170, pp. 167–175, 2024.

[35] V. B. Gowda, G. M. Thimmaiah, M. Jaishankar, and C. Y. Lokkondra, "Background-foreground segmentation using Multi-scale Attention Net (MA-Net): A deep learning approach," Rev. Intell. Artif., vol. 37, no. 3, pp. 557–565, 2023.

[36] V. B. Gowda, M. G. Krishna, and J. Megha, "Dynamic Background Modeling and Foreground Detection using Orthogonal Projection onto the Subspace of Moving Objects," in Proc. IC3, 2023, pp. 171–176.

[37] V. B. Gowda, M. T. Gopalakrishna, J. Megha, and S. Mohankumar, "Background initialization in video data using singular value decomposition and robust principal component analysis," Int. J. Comput. Appl., vol. 45, no. 9, pp. 600–609, 2023.

[38] D. Sumathi, B. Melinamath, and R. Goyal, "Iov Traffic Prediction Utilizing Bidirectional Memory and Spatiotemporal Constraints with Local Search and NonLinear Analysis," Journal of Computational Analysis & Applications, vol. 33, no. 2, 2024.

[39] D. Sumathi, A. Singh, A. Sinha, D. Aditya, and M. R. KF, "The Deepfake Dilemma: Enhancing Deepfake Detection with Vision Transformers," in 2025 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics, Jan. 2025, pp. 1-7.

[40] A. K. Joshi and S. B. Kulkarni, "Flow analysis of vehicles on a lane using deep learning techniques," J. Adv. Inf. Technol., vol. 14, no. 6, pp. 1354–1364, 2023.

[41] A. K. Joshi, V. Shirol, S. Jogar, P. Naik, and A. Yaligar, "Credit card fraud detection using machine learning techniques," Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol., vol. 6, no. 3, pp. 436–442, 2020.

[42] A. K. Joshi and S. B. Kulkarni, "Multi-modal information fusion for localization of emergency vehicles," Int. J. Image Graph., vol. 24, no. 1, Art. no. 2550050, 2024.

[43] A. K. Joshi and S. B. Kulkarni, "Multimodal deep learning information fusion for fine-grained traffic state estimation and intelligent traffic control," Int. J. Intell. Syst. Appl. Eng., vol. 11, no. 3, pp. 1020–1029, 2023.

[44] A. Vahora and K. Pandya, 'A Low-profile 4-element Circularly Polarized Hexagonal DRA Array for Triple-band Wireless Applications', Advanced Electromagnetics, vol. 11, no. 4, pp. 90–97, 2022.

[45] A. Vahora and M. Munsuri, 'Smart Embedded System for Physiological Monitoring Using Machine Learning and Sensor Fusion', Journal of Neonatal Surgery, vol. 14, no. 19s, pp. 694–703, 2025.

[46] M. Fafolawala, Y. Mehta, and A. Vahora, 'Agricultural Drones: Transforming Farming Practices with Advanced Technology', International Journal Of Latest Technology In Engineering,Management & Applied Science, vol. 14, no. 4, pp. 877–882, 2025.

[47] A. Vahora, M. Fafolawala, and Y. Mehta, 'Federated Learning-Enabled Air Quality Monitoring System for Safe Driving in IoT-Integrated Vehicles', International Journal of Environmental Sciences, vol. 11, no. 4s, pp. 715–723, 2025.

[48] V. S. A. Anala, A. R. Pothu, and S. Chintapalli, "Enhancing Preventive Healthcare with Wearable Health Technology for Early Intervention," FMDB Transactions on Sustainable Health Science Letters., vol.2, no.4, pp. 211–220, 2024.

[49] V. S. A. Anala and S. Chintapalli, “Scalable Data Partitioning Strategies for Efficient Query Optimization in Cloud Data Warehouses,” FMDB Transactions on Sustainable Computer Letters., vol. 2, no. 4, pp. 195–206, 2024.

[50] N. Ansari, G. Singh, R. Singh and Sheetal, "Innovative herbal tea formulation using Holarrhena antidysenterica, Emblica officinalis, and Stevia: Nutritional and phytochemical analysis," J. Neonatal Surg., vol. 14, no. 6, pp. 381-389, 2025.

[51] S. Bala, G. Singh and M. Kaur, "Mindfulness of functional foods in cancer prevention and health promotion: A comprehensive review," Rev. Electron. De Vet., vol. 25, no. 1, pp. 1181-1187, 2024.

[52] S. Bala, G. Singh, R. Arora and Devanshika, "Impact of caffeine consumption on stress management and stamina among university students," Rev. Electron. De Vet., vol. 25, no. 2, pp. 253-259, 2024.

[53] G. Singh and M. Khatana, "Assessment regarding the efficacy of intermittent fasting," Int. J. Res. Anal. Rev., vol. 9, no. 1, pp. 2349-5138, 2022.

[54] G. Singh et al., "Analyze the effects of prebiotics on the immunity of human beings through various clinical studies," Jundishapur J. Microbiol., vol. 15, no. 1, pp. 1167-1177, 2022.

[55] G. Singh et al., "Comprehensive look of renal calculi in kidneys: A review," NeuroQuantology, vol. 20, no. 5, pp. 4404-4412, 2022.

[56] Md H. Rahman, T. Islam, Md E. Hossen, Md E. Chowdhury, R. Hayat, Md S. S. Shovon, M. Alamgir, S. Akter, R. Chowdhury, and A. R. Sunny, "Machine Learning in Healthcare: From Diagnostics to Personalized Medicine and Predictive Analytics," J. Angiother., vol. 8, no. 12, pp. 1–8, 2024.

[57] R. Chowdhury, Md A. H. Fahad, S. M. S. Alam, M. I. Tusher, Md N. U. Rana, E. Ahmed, S. S. Akhi, and Md R. H. Mahin, "Database Management in the Era of Big Data: Trends, Challenges, and Breakthroughs," Pathfinder Res., vol. 1, no. 1, p. 15, 2020.

[58] Md R. H. Mahin, E. Ahmed, S. S. Akhi, Md A. H. Fahad, M. I. Tusher, R. Chowdhury, and Md N. U. Rana, "Advancements and Challenges in Software Engineering and Project Management: A 2021 Perspective," Pathfinder Res., vol. 2, no. 1, p. 15, 2021.

[59] Md A. H. Fahad and R. Chowdhury, "Evolution and Future Trends in Web Development: A Comprehensive Review," Pathfinder Res., vol. 3, no. 1, p. 13, 2022.

[60] S. N. Akhter, R. Kumari, and A. Kumar, “Fertility booster effect of Asparagus recemosus against arsenic induced reproductive toxicity in Charles Foster rats,” J. Adv. Zool., vol. 45, no. 5, 2024.

[61] Z. Hashmi, R. Kumari, and A. Kumar, “Antidote effect of Bacopa Koneru against arsenic induced toxicity in rats,” J. Adv. Zool., vol. 45, no. 5, 2024.

[62] Z. Hashmi, R. Kumari, and A. Kumar, “Phytoremedial effect of Ocimum sanctum against arsenic induced toxicity in Charles Foster rats,” J. Adv. Zool., vol. 45, no. 5, 2024.

[63] B. Kumari, P. Das, and R. Kumari, “Accelerated processing of solitary and clustered abasic site DNA damage lesion by APE1 in the presence of aqueous extract of Ganoderma lucidum,” J. Biosci., vol. 41, pp. 265–275, 2016.

[64] R. Kumari, R. K. Singh, N. Kumar, and R. Kumari, “Preparation of superfine Bael leaf nanopowder, physical properties measurements and its antimicrobial activities,” Egypt. Chem. Bull., vol. 12, no. 4, pp. 284–297, 2023.

[65] M. K. Sinha, R. Kumari, and A. Kumar, “Ameliorative effect of Ganoderma lucidum on sodium arsenite induced toxicity in Charles Foster rats,” J. Adv. Zool., vol. 45, no. 5, 2024.

[66] V. Rajavel, "Integrating power-saving techniques into design for testability of semiconductors for power-efficient testing," The American Journal of Engineering and Technology, vol. 7, no. 3, pp. 243–251, 2025.

[67] V. Rajavel, "Novel machine learning approach for defect detection in DFT processes," ASRJETS-Journal, vol. 101, no. 1, pp. 325–334, Apr. 2025. [Online].

[68] V. Rajavel, "Optimizing semiconductor testing: Leveraging stuck-at fault models for efficient fault coverage," Int. J. Latest Eng. Manag. Res. (IJLEMR), vol. 10, no. 2, pp. 69–76, Feb. 2025.

[69] Md S. Miah and Md S. Islam, "Big Data Analytics Architectural Data Cut off Tactics for Cyber Security and Its Implication in Digital forensic," in Proc. 2022 Int. Conf. Futuristic Technol. (INCOFT), Belgaum, India, 2022, pp. 1–6.

[70] M. A. Obaida, Md S. Miah, and Md. A. Horaira, “Random Early Discard (RED-AQM) Performance Analysis in Terms of TCP Variants and Network Parameters: Instability in High-Bandwidth-Delay Network,” Int. J. Comput. Appl., vol. 27, no. 8, pp. 40–44, Aug. 2011.

[71] A. Srivastava, “Use of Python in Data Science, Data Integration and Data Engineer,” Int. J. Sci. Res. Eng. Manag., vol. 8, no. 7, 2024.

[72] A. Srivastava, “AI in Healthcare and its Future,” J. Artif. Intell. Cloud Comput., vol. 1, no. 1, pp. 1–2, Mar. 2022.

[73] A. Srivastava, “Cloud Replacing Traditional Database,” Int. J. Multidiscip. Res., vol. 7, no. 2, pp. 1–2, Mar.–Apr. 2025.

[74] A. Srivastava, “Data Transformation Normalization to Denormalization in Cloud,” Int. J. Core Eng. Manag., vol. 6, no. 7, pp. 249–252, 2020.

[75] A. Srivastava, “Impact of AI/ML on Job Market and Skills Set and Health Industry,” ESP J. Eng. Technol. Adv., vol. 4, no. 3, pp. 122–126, 2024.

[76] G. Kashyap, "Neural Architecture Search (NAS): Exploring the Trade-Offs In Automated Model Design and Its Impact on Deep Learning Performance," Int. J. Innov. Res. Eng. Multidiscip. Phys. Sci., vol. 13, no. 2, pp. 1–12, Mar.–Apr. 2025.

[77] G. Kashyap, "Large Language Models and Their Ethical Implications: The role of models like GPT and BERT in shaping future AI applications and their risks," Int. J. Innov. Res. Creat. Technol., vol. 6, no. 6, pp. 1–5, Dec. 2020.

[78] G. Kashyap, "AI for Epidemiology: Using AI to Predict and Track the Spread of Diseases like COVID-19," Int. J. Innov. Res. Multidiscip. Field, vol. 3, no. 6, pp. 1–10, Nov. 2021.

[79] G. Kashyap, "AI for Threat Detection and Mitigation: Using AI to Identify and Respond to Cybersecurity Threats in Real-Time," Int. J. Sci. Res. Eng. Manag. Sci., vol. 6, no. 6, pp. 1–5, Nov. 2024.

[80] G. Kashyap, "AI for Information Retrieval: Advancements in Search Engines and Chatbots through Deep Learning-Based Query Understanding," Int. J. Innov. Res. Creat. Technol., vol. 7, no. 1, pp. 1–7, Jan. 2021.

[81] G. Kashyap, "Multilingual NLP: Techniques for Creating Models that Understand and Generate Multiple Languages with Minimal Resources," Int. J. Sci. Res. Eng. Manag. Sci., vol. 6, no. 12, pp. 1–5, Dec. 2024.

[82] G. Singh, S. Bala and S. Singh, "Nutraceuticals miraculously alter human genomics: A review on nutrigenomics," Afr. J. Biol. Sci., vol. 6, no. 6, pp. 599-699, 2024.

[83] G. Singh, S. Bala, M. Kaur and S. Phagna, "Exploring public awareness and attitudes towards dietary supplements," Afr. J. Biol. Sci., vol. 6, no. 6, pp. 7288-7299, 2024.

[84] G. Singh, J. Bharti and C. Dua, "Mindfulness of nutritional knowledge and food hygiene practices on the health among young adults," Afr. J. Biol. Sci., vol. 6, no. 6, pp. 5813-5818, 2024.

[85] G. Singh and G. K. Kochar, "Zinc content of commonly consumed foods of Kurukshetra district of Haryana," Food Sci. Res. J., vol. 1, no. 2, pp. 94-98, 2010.

[86] G. Singh et al., "Liver cirrhosis: The struggling liver," Int. J. Health Sci., vol. 6, no. 1, pp. 5547-5559, 2022.