Date of publication SEP-30, 2024, date of current version SEP-09, 2024.
www.computingonline.net / computing@computingonline.net

Print ISSN 1727-6209
Online ISSN 2312-5381
DOI 10.47839/ijc.23.3.XXXX

Efficient Deep Learning Methods for Detecting Road Accidents by Analyzing Traffic Accident Images

P.C. SHERIMON[1], VINU SHERIMON[2], JOHNSYMOL JOY[3], AMBILY MERLIN KURUVILLA[3], G. ARUNDAS[3]
[1]Faculty of Computer Studies, Arab Open University, Muscat, Sultanate of Oman
[2]College of Computing and Information Sciences, University of Technology and Applied Sciences, Muscat, Sultanate of Oman
[3]P.G. Department of Computer Applications and Artificial Intelligence, Saintgits College of Applied Sciences, Kerala, India

Corresponding author: Vinu Sherimon (e-mail: vinusherimon@yahoo.com).

The research leading to these results has received funding from the Research Council (TRC) of the Sultanate of Oman under the Block Funding Program BFP/RGP/ICT/22/143.

ABSTRACT Speed is one of the major factors in car crashes. Many lives could have been saved if emergency services had been alerted to the disaster and arrived in time. For the sake of protecting valuable human lives, an effective automatic accident detection system with prompt reporting of the accident scene to emergency services is essential. Therefore, this research proposes some effective Deep Learning techniques that properly recognize the incidence of accidents. The paper introduces two different techniques for image classification, with a particular focus on distinguishing between accident and non-accident images. The dataset used for the proposed model is taken from Kaggle, which is a collection of CCTV images. The first approach is a hybridized TL-ML method that employs transfer learning techniques that use different pre-trained versions of convolutional neural networks to extract features from image datasets. These extracted features are then fed into various machine learning classifiers to categorize the images as either Accident or Non-accident. To make the final decision, a voting classifier is utilized to choose the best classification outcome from the set of previously employed machine learning classifiers. In the second method, a modified Convolutional Neural Network (CNN) called SpinalNet is adopted. The performance of these models was evaluated by comparing them with each other and with a customized CNN base model. SpinalNet consistently surpassed the other models in terms of Precision, Recall, F1-Score, and Accuracy, demonstrating its outstanding capabilities.

KEYWORDS CNN; Deep Learning; VGG16; Inception v3; Feature Extraction; Machine Learning; SpinalNet

I. INTRODUCTION

ROAD traffic injuries are a major public health issue around the world, accounting for around 1.3 million deaths each year [1]. Children and young adults aged 5 to 29 are especially vulnerable, as traffic accidents are the main cause of death in this age group [1]. Furthermore, despite having just 60% of the world's automobiles, low- and middle-income countries are disproportionately afflicted by road traffic accidents, accounting for 93% of all road fatalities [1]. The economic impact of traffic accidents is also significant, with most countries losing 3% of their GDP because of traffic accidents [1].

Road traffic accidents have a huge societal impact, resulting in a wide range of social, economic, and personal losses [1]. Road accidents frequently result in fatalities, resulting in irreplaceable human loss. Accidents can sometimes cause serious injuries, resulting in disability and long-term health issues. Road accidents have a huge economic impact as well, as they result in lost productivity and increased healthcare costs [2]. Accidents can also result in expensive fixes to automobiles, infrastructure, and other property. Accidents can cause traffic congestion, resulting in delays and longer travel times [3, 4]. This can harm the economy by making it more difficult for people to get to work, school, and other important regions. Individuals and communities suffer substantial emotional consequences because of road accidents, including anxiety, stress, and trauma [5]. This can have an impact not only on the persons involved in the disaster, but also on their families, friends, and anyone who sees the aftermath. Given these implications, detecting road accidents is crucial to lowering the harmful consequences of accidents. Early detection of accidents allows emergency services to be contacted immediately, potentially reducing the severity of injuries and the economic and emotional consequences of accidents. Deep learning algorithms like SpinalNet have the potential to enhance the accuracy and efficiency of traffic accident detection, hence making roads safer for everyone. The goal of this research is to develop an efficient accident detection system utilizing deep learning technologies for prompt emergency intervention.

The rest of the paper is organized as follows: Section 2 discusses the prominent technologies used in road accident detection. Deep learning technologies are described in Section 3. Literature review is given in Section 4. Section 5 includes the proposed methodology. Model generation using augmentation techniques and SpinalNet is given in Section 6. Section 7 describes the results and discussion. Conclusion & Future is given in Section 8 followed by References.

II. PROMINENT TECHNOLOGIES USED IN ROAD ACCIDENT DETECTION

Several technologies, including deep learning, computer vision, sensors, machine learning, IoT, data analytics, GIS, and mobile/wireless technologies, are used to detect road accidents. To identify accident trends, deep learning systems evaluate sensor data or video footage. To find anomalies and pinpoint accidents, computer vision systems analyze photos and videos. GPS and accelerometer sensors offer information on vehicle motions for accident detection. From previous accident data, machine learning systems discover trends. IoT makes it possible for infrastructure, sensors, and vehicles to exchange data in real-time. Accident data is spatially analyzed using data analytics and GIS to pinpoint accident-prone locations. Real-time data collection is done using wireless and mobile technologies for analysis and warnings. The potential for detecting and responding to traffic accidents has been improved by these technologies.

Several research studies have been carried out to address the issue of road accidents, with a special emphasis on increasing road safety through the development of modern technologies. Intelligent transportation systems based on fog robotics integrate IoT devices and FC centers for effective data processing. The proposed system [6], which emphasizes self-directed transportation, makes use of wireless decentralized sensors and intelligent speed assistance powered by AI. It seeks to maximize energy efficiency, traffic management, and road safety. Recent studies, for example, have investigated the use of computer vision and deep learning approaches to improve accident detection and prevention [7–9]. Other research has examined the usefulness of intelligent transportation systems (ITS) in lowering the number and severity of traffic accidents [10]. Planning for road safety must consider the causes and consequences of incidents.

This study [11] examined driving practices and accident outcomes in Al-Ahsa city, Saudi Arabia, using a Bayesian belief network (BBN) model. When speeding and brake failure are considered simultaneously, the findings showed a considerable increase in the chance of a collision. The complicated relationships between driving behavior and accident causes were effectively explored by the BBN model. High-resolution driving behavior data from 303 drivers was gathered by this study [12] employing smartphone sensors. The study combined the data with information on traffic and road geometry, concentrating on road segments and junctions. The investigation showed that traffic features, such as flow, occupancy, and speed, had a greater impact on the frequency of unpleasant driving episodes than factors like road layout and driver behavior. In nations where accurate accident data is still missing, road safety audits play a vital role in improving road safety. This study [13] focuses on evaluating the RSA of a portion of National Highway 326 with the goal of identifying accident-prone locations and examining the effects of traffic patterns and road layout.

Road accident detection has been successfully accomplished using machine learning and deep learning technology. They are useful tools for assessing complicated and dynamic traffic circumstances due to their capacity to learn from big datasets, identify patterns, and generate accurate predictions. In this research, we propose a novel deep learning technique using SpinalNet for predicting road accidents.

III. MACHINE LEARNING TECHNOLOGIES

Machine learning technologies play a crucial role in today's data-driven world, empowering computers to learn patterns from data and make informed decisions. A classifier is a machine learning model that has been trained to identify a given input sample's class or category. Among various machine learning techniques, Random Forest, Extra Trees, Decision Trees, and Logistic Regression stand out as widely used and powerful algorithms.

Random forest classifier [14] - A random forest is an ensemble learning technique for classification, regression, and other tasks. It works by building many decision trees during the training phase, then producing the class that represents the mean of the classes (for classification) or the mean prediction (for regression) of the individual trees. It is a kind of machine learning method that is utilized for both regression and classification problems. The name's "forest" and "random" components allude to the several decision trees that make up the model, respectively. The trees are built using the random subspace method and random sampling of data points (with replacement).

Decision tree classifier [15] - An internal node represents a feature (or attribute), a branch denotes a classification algorithm, and each leaf node represents the result in a decision tree, which resembles a flowchart. The root node in a decision tree is the first node from the top. From the values of the input features, it learns to divide the data into subsets. The tree uses a decision rule at each internal node to determine whether to divide the data based on one of the input features. Until a stopping criterion, such as a maximum tree depth or at least number of samples in a leaf node, is satisfied, this process iterates for each internal node. The path from the root to a leaf node indicates a classification determination rule, and the leaf nodes themselves reflect the class labels at the conclusion. Understanding, interpreting, and visualizing it is easy. Decision trees are frequently employed in a variety of real-world settings, including banking, healthcare, and marketing.

Extra tree classifier [16] - The ensemble learning technique Extra Trees (Extremely Randomized Trees) is used for classification and regression tasks. It is a version of the Random Forest algorithm in which random splits are used to construct the trees rather than the best splits obtained through feature selection. As a result, the trees become more diverse, which enhances the model's overall performance. The Extra Trees method is renowned for its resilience to overfitting and capacity to handle high-dimensional data. For many machine learning tasks, including image classification, natural language processing, and bioinformatics, it is a preferred option.

Logistic regression [17] - A statistical technique called logistic regression is used to forecast events that can only have one of two possible outcomes, such as true or false. It is a kind of generalized linear model that simulates a binary dependent variable using a logistic function. Finding the optimal set of parameters for the model's independent variables is the aim of logistic regression to make the projected chance of the binary outcome as accurate as possible. Both linear and non-linear decision boundaries can be handled by logistic regression.

IV. DEEP LEARNING TECHNOLOGIES

The development of algorithms and statistical models that allow computers to learn from and make predictions or choices without being expressly programmed to do so is the field of machine learning, a branch of artificial intelligence. Machine learning comes in a variety of forms, such as reinforcement learning, unsupervised learning, and supervised learning. Predictive analytics, computer vision, and natural language processing are just a few of the many applications where machine learning is applied. A labeled dataset is used to train the algorithm in supervised learning, a type of machine learning. The dataset is made up of input-output pairs, where the input consists of a set of features (sometimes referred to as predictors or attributes) and the output is the associated label or target variable. There are two primary types of supervised learning: classification and regression. The aim of supervised learning is to establish a mapping from inputs to outputs so that the algorithm can make precise predictions on fresh, unseen data. While the objective of regression is to predict a continuous value, the objective of classification is to predict a categorical label. Support vector machines, logistic regression, and decision trees are a few examples of supervised learning algorithms. Predictive analytics, natural language processing, and picture classification are just a few of the many applications that make use of these algorithms.

A. DNN

A kind of machine learning known as "deep learning" uses neural networks with several layers to learn data representations. Deep neural networks (DNNs), which are a type of neural network, have the capacity to learn intricate patterns and features from vast volumes of data. It is modeled after the structure and operation of the human brain. Artificial neurons are layers of interconnected nodes that process and transfer information. The artificial neuron is the fundamental unit of a neural network. It accepts inputs, processes them, and outputs the results. The output of the neuron is produced after the inputs have been multiplied by a set of weights and processed through an activation function. To effectively predict or classify the output given a set of inputs, a neural network must be trained, which entails changing the weights of the artificial neurons. Using a labeled dataset, input-output pairs are given to the network, and the weights are changed to reduce the difference between the predictions made by the network and the actual output. With their capacity to learn from big and complicated datasets and generalize, neural networks have tremendous potential.

Deep learning models may develop more abstract representations of the data thanks to its hierarchical structure, which is advantageous for tasks like image identification and natural language processing. Feedforward neural networks, convolutional neural networks, and recurrent neural networks are a few examples of deep learning architectures. These designs are frequently employed in a wide range of applications, including speech recognition, natural language processing, and computer vision. Due to the accessibility of enormous amounts of data and the advancement of more potent hardware, like GPUs, deep learning has gained popularity in recent years. These developments have made it possible for deep learning models to function at the cutting edge across a variety of tasks.

Although DNNs have many benefits, there are a few problems that have been found and are still being worked on in their development and use. The fact that DNNs often need a lot of input features is one problem. While adding additional parameters may improve prediction accuracy, doing so also makes the network more complex and computationally intensive. The ideal thickness of the first hidden layer must also be established. While a large first hidden layer results in a huge increase in the number of weights, making training more difficult, a tiny first hidden layer may not transmit all input features efficiently. Another drawback of conventional DNNs is the vanishing gradient issue. The gradient, which is essential for updating weights during training, decreases as the number of layers rises because it propagates back to neurons close to the inputs. As a result of this gradient vanishing phenomena, previous layers of deep networks acquire very little training information. These issues are being actively addressed by academics and researchers, who are also creating methods that reduce the challenges involved with DNNs.

B. CNN

CNN [18] are a type of deep learning network architecture that automatically extracts features from input rather than learning from it manually. An example of a neural network architecture that is frequently used for image and video processing applications is CNN (Convolutional Neural Network) layers. CNN is frequently employed for image recognition and other visual tasks. It is made up of several layers of neurons, each of which is linked to a discrete area of the input image. Each layer's neurons are arranged into "feature maps" that are used to identify various features in the image. CNNs are frequently employed in applications like object identification and image classification because they are particularly good at identifying patterns and characteristics that are present in various sections of an image.

V. RELATED STUDIES

Numerous literature papers were considered and analyzed to recognize the merits and demerits of various ideas related to accident detection. An ensemble deep learning [19] for car crash detection that used both video and audio data from the dashboard cameras were checked. They used CNN and GRU based classifiers i.e., Convolutional Recurrent Neural Network (CRNN) for detection. CNN is used to extract features from video files, but it causes memory issues to prevent them from using video generators. The proposed model produced an ROC-AUC value of 98.60.

The research [20] includes the use of mobile applications, vehicle ad hoc networks, GSM and GPS technologies, and smartphone-based accident detection. The gyroscope and vibration sensors detect an accident, and the GSM module quickly sends a message to the emergency contact numbers along with the accident's location. Another study [21] is to develop a Real Time Traffic Accident Detection System (RTTADS) using RFID and Wireless Sensor Network (WSN) technologies. This paper describes the hardware prototype configuration for RTTADS, the employed algorithms, and the benefits and drawbacks of the complete system.

The YOLOv5 deep learning system is used in [22] to identify automobiles in real-time CCTV surveillance footage. The system's main goal is to create a model that can identify different vehicle classes using a unique dataset. The dataset consists of 1000 photos taken under diverse situations, including rain, poor visibility, brightness, and weather. They obtained an accuracy of 98%. In [23] two models were used, which are; RNN for image feature extraction and for detection they used dense ann. The accuracy obtained is 0.98. The solution is restricted to vehicle collection i.e., by excluding motor vehicles, bicycles. Also, the model had errors in determining accident segments with low illumination i.e., at nighttime or low resolution. To execute this experiment, it requires large data with clear data. The accuracy of this model is 98%.

Research [24] employed neighborhood component analysis, a supervised learning approach based on K nearest neighbors, and a partial dependence plot. and individual conditional expectation-based feature selection approach for examining the major causes of the severity of traffic accidents. Following the application of those algorithms to the data set, they plotted the significance of factors for visualizing our results and drew up a table to display the numerical output of each algorithm individually. Finally, we employed support vector machine classification with all 15 variables and the eight that were found to be most important for validating and testing our findings. By contrasting the outcomes of two support vector machine classification models, we have demonstrated the validity and properly supported our study. The accuracy of both models explained in this model is 89.9%.

When detecting accidents, choosing the appropriate camera perspective is quite important. The research in [25] outlines a deep learning architecture to examine the accident events as they were captured from various angles. Then they start by estimating feature similarity in videos taken from various angles. The video samples were then split into groups with high and low feature similarity. Then, using two-branch DCNNs to extract spatio-temporal features from each group, they fuse the features using a rank-based weighted average pooling technique before classifying the results.

Research work [26] devised a brand-new technique that instantly detects accidents by using objects and their locations. They did high-level post processing in addition to localizing the accident occurrences in videos to depict the gravity and context of an accident. To extract object interactions, they first separated an input video into pre-accident, accident, and post-accident periods. A refining process is then used to filter these interaction proposals. Then they use an iterative training process to categorize regular encounters and accidents. Using heat maps, they additionally highlight the damaged area. To measure the context and severity of an accident, they provide high-level linguistic descriptions. The AUC values of UCF Crime and CADP are 69.70% and 72.59% respectively.

The goal of the research work [27] is to completely automate the operational reconstruction of an accident site to guarantee the highest degree of accuracy when measuring the distances between items' relative locations. A road accident site is first marked by the operator, after which the UAV scans and gathers data on the site. They created a three-dimensional accident scenario. Then, using the deep learning model SWideRNet with Axial Attention, items of interest on the three-dimensional scene are segmented. An image transformation approach and marked-up data are used to build a two-dimensional road accident scheme. The segmented items' relative locations inside the scheme, between which the distance is determined, are included. They evaluated the precision of the segmentation of the reconstructed items using the Intersection over Union (IoU) measure. To assess the precision of the automatic distance measurement, they employed the Mean Absolute Error. The accuracy of their proposed model is 77.1%.

To improve road safety and security, article [28] suggests a deep learning architecture called ConvLSTM for driver drowsiness detection. On the Yawn Eye Dataset with 4 classes (Closed, Open, No Yawn, Yawn), the suggested ConvLSTM model demonstrated accuracy of 99.44%, while on the MRL Dataset with 37 classes, it demonstrated accuracy of 90.12%. (s0001, s0002, s00037). This work suggests a method for automatically choosing regions of interest using stacked spatiotemporal convolution and long short-term memory. Drowsiness detection neural network (ConvLSTM) for an in-car security and surveillance system. Classifiers in the Haar Cascade on the human face to choose the area-of-interest. To extract spatio-temporal information from the chosen region-of-interest and to foretell the driver's level of tiredness, a ConvLSTM model is implemented.

Paper [29] proposes the idea of object detection using the YOLO algorithm. To run at a depth of 12 layers and be trained on a total of more than 80 different classes, the YOLO algorithm demands not only a lot of data but also very substantial processing resources. They trained the YOLO algorithm in as few classes as possible to prevent overloading and to reduce the resource consumption needed for processing. The classes selected for training are as follows: Trees, a person, a bus, a bicycle, a truck, a car, a road sign.

In paper [30] Convolutional neural network topologies were examined, and it was discovered that transfer learning was the most effective method. The most effective foundation models for transfer learning were EfficientNetB1 and MobileNetV2. The first for its accuracy in prediction, and the second for its magnitude and speed of execution. Images from Finnish road surveillance cameras that are made available as open data every ten minutes were used as a case study. The result was discovered to have a mean average precision of 0.89 and a Matthews Correlation (MCC) of 0.77 for the solution trained using EfficientNetB1 as the base model. The MobileNetV2-based system has an MCC of 0.71 and a mAP of 0.88.

VI. PROPOSED METHODOLOGY

The paper presents three methods for image classification, specifically focused on distinguishing between accident and non-accident images. The first two methods use a transfer learning approach to extract features from image datasets, then use different Machine learning classifiers to categorize images as Accident or Non-accident, and finally use a voting classifier to select the best classification outcome from the previously used set of ML classifiers. We called them Hybrid TL-ML (Transfer Learning - Machine Learning) models. The advantage of this hybrid approach lies in leveraging the powerful feature extraction capabilities of transfer learning along with the generalization abilities of machine learning classifiers for accurate classification. The third method adopts a modified CNN known as SpinalNet. This CNN architecture has been customized or optimized for the specific task of image classification. By utilizing SpinalNet, the authors aim to achieve precise classification results while capitalizing on the robustness and efficacy of CNNs in handling image-related tasks. Performance of these three models were compared with a customized CNN base model.

The dataset is a collection of CCTV footage of road accidents which is taken from Kaggle [31]. The dataset is divided into three subsets: training, testing, and validation. Each subset consists of two classes: Accident and Non-accident. In the training subset, there are 369 accident images and 421 non-accident images. The testing subset contains 47 accident images and 53 non-accident images. Finally, the validation subset includes 46 accident images and 52 non-accident images. Different models were proposed for training the system.

A. CUSTOMIZED BASE MODEL USING CNN

First, we use CNN as a base model which contains an input layer with input shape (250,250,3), 3 convolutional layers and corresponding max-pooling layers, flatten layers and dense layers.

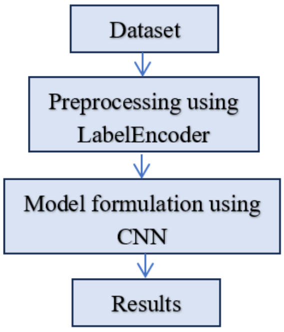

Figure 1. CNN Model Architecture.

Initially, the collected dataset is preprocessed using LabelEncoder and formulated the CNN model [32] for training and evaluating the data. LabelEncoder is a preprocessing technique commonly used for encoding categorical labels into numeric representations. It is frequently utilized to translate target labels or categorical information into numerical values that machine learning algorithms can comprehend. Fig. 1 shows CNN Model architecture diagram.

The pooling layer is used to reduce the dimensions of the feature maps. As a result, it decreases the quantity of network computation and the number of parameters to learn. The feature map created by a convolution layer's feature pooling layer summarizes the features that are present in a specific area. The flatten layer reduces the output of the convolutional layer, which is 118x118x10, to a single, one-dimensional vector that can be utilized as the input for a dense layer [33].

The convolutional layer, which applies a series of filters to the input data to extract features, is the fundamental component of CNN. These filters are made to identify specific patterns in the input data and are often modest in size (e.g., 3x3 or 5x5 pixels) (e.g., edges, textures, shapes). The output of a convolutional layer is often down sampled using a pooling layer after being passed through a non-linear activation function (like ReLU). The process is then performed with a new convolutional layer using the pooled feature maps. To categorize the characteristics discovered by the convolutional layers, fully connected layers are typically added at the very end of the CNN architecture. A set of scores, one for each class in the dataset, is the output of the last fully connected layer. This model utilizes the Adam optimizer and employs sparse categorical cross-entropy for loss calculation. The dataset is divided with 80% for training, 10% for validation, and another 10% for testing. Here for enhancing the performance of the CNN, additional layer types, including dropout layers and normalization layers (Batch normalization) are added.

B. HYBRIDIZED TL-ML MODELS

Transfer learning (TL) is the term used in machine learning to describe using a previously trained model on a different task [34]. In transfer learning, a machine uses the information gained from a previous work to improve prediction about a new task. TL allows different domains, tasks, and distributions to be used for training and testing.

In this work, feature extraction is carried out using two methodologies VGG16 [35] and InceptionV3 [36]. VGG16, with its deep architecture and pre-trained weights, is well-suited for transfer learning in different computer vision applications. Feature Extraction is the process of transforming raw data into numerical information without making any changes in the original dataset. The extracted features are fed into machine learning classifiers for improving accuracy. The problem, the quantity and caliber of the training data, and the required level of accuracy all influence the classifier that is selected. Different classifiers are better suited to different sorts of data or tasks, and some are easier to understand or more computationally efficient than others. But here, we used a voting classifier to combine the prediction.

A voting classifier is a machine learning model that trains on a large ensemble of models and predicts an output (class) based on the highest probability of the chosen class being the outcome. It simply accumulates the results of each classifier that has been passed into voting classifier and predicts the output class based on the highest majority of votes. Instead of developing separate specialized models and determining their correctness, we develop a single model that trains on these models and predicts output based on their aggregate majority of voting for each output class.

a) Model 1 – Hybridized TL-ML model using VGG16

The ImageNet dataset, which is frequently used for image classification and other computer vision tasks, was used to train the convolutional neural network model VGG-16. The model was created by the Visual Geometry Group at the University of Oxford, which is indicated by the initials "VGG" in the name. The number 16 indicates that there are 16 layers in the model. The VGG16 architecture is renowned for its use of small convolutional filters and deep architectures, and it consists of several convolutional layers and fully linked layers.

In the first model VGG16 is used for feature extraction and finally fed the extracted features to different [37] ML classifiers. First, the VGG16 model [38] from Tensor flow Keras is imported here. Both the pre-process input and image modules are imported to scale the pixel values for the VGG16 model suitably. The image module pre-processes the image object. Then the pre-trained weights for the ImageNet dataset are input into the VGG16 model [38]. A sequence of dense (or completely linked) layers are placed after each convolutional layer in the VGG16 model. The final dense layers can be chosen or deselected using Include top. False means that while loading the model, the last dense layers are not included. The network's remaining layers are regarded as the model's classification portion, with the input layer through the final max pooling layer being considered its feature extraction portion. Figure 2 illustrates the architecture of the feature extraction using VGG16. The extracted feature is fed into four classifiers: random forest, logistic regression, extra tree, and decision tree. Then, the voting classifier combines the predictions of multiple individual classifiers (e.g., Random Forest, Logistic Regression, Extra tree, and Decision Tree) and assigns the final prediction based on majority voting. This approach can help exploit the strengths of different feature extraction techniques and classifiers, providing better generalization.

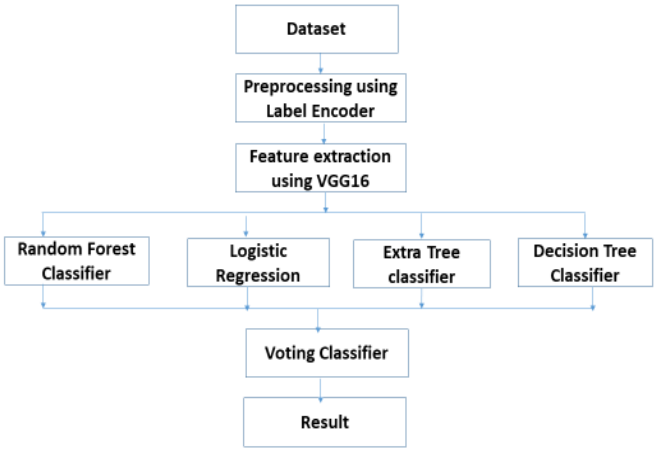

Figure 2. Feature Extraction using VGG16.

b) Model 2 – Hybridized Tl-Ml Model Using Inception V3

A convolutional neural network design for image classification tasks is called Inception v3 [39] [40]. Inception modules, which are building components that combine convolutional and pooling layers to identify features at various scales, are a hallmark of the design. The Inception v3 design is more complex than the Inception v2 architecture and was trained using the ImageNet dataset, where it attained the best performance available at the time. The model is frequently employed for object identification, image classification, and other computer vision applications.

To extract valuable features from fresh samples, we can utilize a model that has already been trained. The pre-trained model is simply added on top of a proposed classifier that will be trained for feature maps that were previously learned for the dataset. The entire model does not have to be retrained. In the second model Inception v3 is used for feature extraction and finally fed the extracted features to random forest, logistic regression, extra tree, and decision tree ML classifiers. Like VGG16 model, here also the voting classifier is used to combine the predictions from multiple classifiers. Figure 3 demonstrates the architecture of feature extraction using Inception v3.

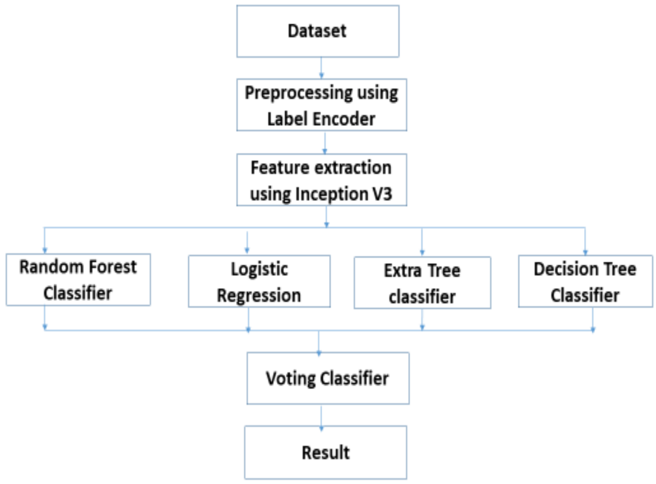

Figure 3. Feature Extraction using Inception v3.

C. MODEL USING SPINALNET

a) Model 3-Spinalnet Model

An artificial neural network called a SpinalNet is fashioned to resemble the form and operation of the spinal cord. The spinal cord oversees sending and receiving signals from the brain to the rest of the body. It also controls some reflexes and simple motor movements. To quickly receive vast amounts of data and perform better, SpinalNet attempts to emulate the human somatosensory system. Figure 4 shows the human somatosensory system, describing how the sensory information from our bodies is received by the spinal cord. The area of the nervous system in humans that oversees processing sensory data from the body is known as the somatosensory system. Information on proprioception, pain, temperature, and pressure is also included (awareness of the position and movement of the body). The somatosensory system includes specific receptors in the skin, muscles, joints, and internal organs as well as pathways in the spinal cord and brain that communicate and process this data. Numbness, tingling, and chronic pain are just a few of the disorders that can result from harm or malfunction to the somatosensory system.

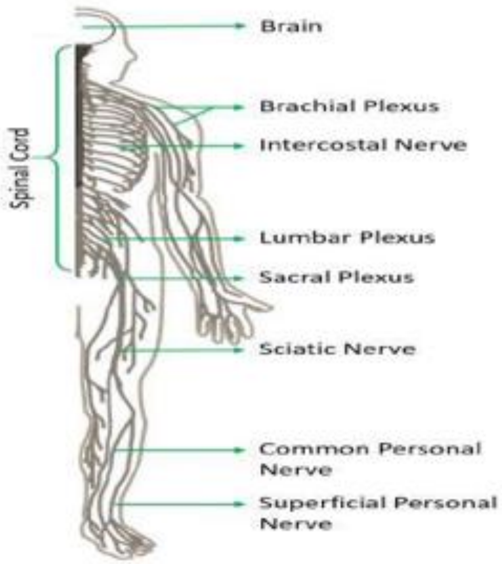

Figure 4. Human somatosensory system [41].

A spinal net is designed to mimic similar actions in a machine, enabling it to react swiftly and autonomously to environmental changes. In fields like robots, prosthetics, and others where real-time control and sensing are crucial, this can be helpful. Spinal nets and convolutional neural networks (CNNs) are both categories of artificial neural networks, although their structures and intended applications differ. A SpinalNet is a special kind of neural network that is based on the design of the spinal cord. Reflexive actions and motor control capabilities of the spinal cord are intended to be mimicked in artificial systems, such as robots or prosthetics. The system can react swiftly to changes in its surroundings because it is designed to process sensor input and control decisions in real-time. In short, CNNs are mainly used for image processing and pattern recognition, while spinal nets are mainly used for real-time control in robotic systems.

Figure 5 shows the input row, the intermediate row, and the output row that make up the proposed neural network. There are several hidden layers in the intermediate row. A piece of the input is sent to each buried layer. The outputs from the preceding layer are also sent to every layer except the first. The weighted outputs of each hidden neuron in the intermediate row are added to the output layer. A SpinalNet [43] can be built and trained for any unspecified number of inputs, intermediate neurons, and outputs.

The SpinalNet [41] can be used as the fully connected or classification layers which supports both traditional learning and transfer learning. The hidden layer architecture in SpinalNet differs from a typical neural network model. In a neural network, the hidden layers pass intermediate outputs to the following layer after receiving inputs from the previous layer. But in SpinalNet, the hidden layer makes it possible for the top layer to receive some of its inputs and outputs. As a result, the hidden layer of neural networks has fewer incoming weights than typical neural networks. Figure 6 shows the architecture of a model using SpinalNet.

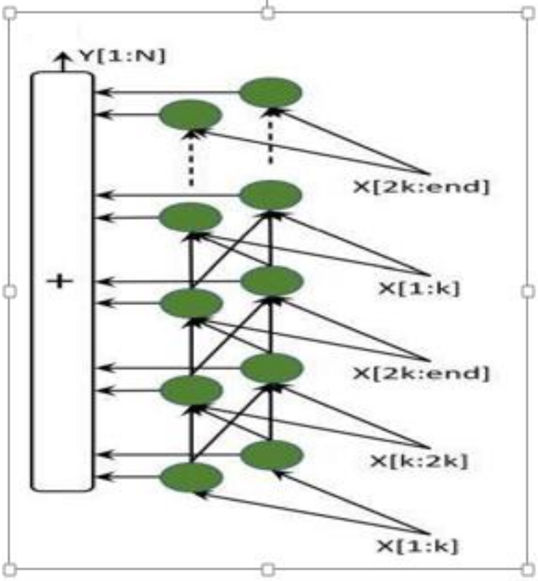

Figure 5. Structure of SpinalNet [41].

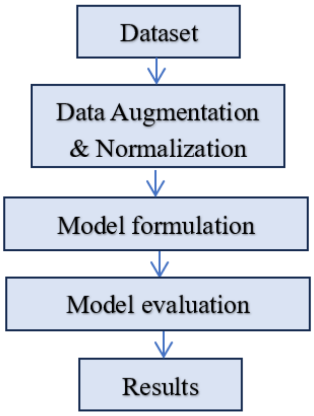

Figure 6. SpinalNet model.

VII. RESULTS AND DISCUSSION

Python was used to develop the proposed models, Hybridized TL-ML models (Model 1 and Model 2) and SpinalNet (Model 3). The performance was assessed and compared with the base model. Precision, Recall, F1-Score, and Accuracy are the variables used to assess the performance. The dataset is a collection of kaggle-sourced CCTV footage of road accidents. The dataset is divided into three subsets: training, testing, and validation. 80% for training, 10 percent for testing, and 10 for validation.

When we want to measure positive cases very precisely, precision is used. It determines what proportion of situations that are expected to be positive turn out to be so. When capturing the most positive cases possible, recall is used. It determines the accuracy with which true positive instances are identified. F1 is typically more helpful than accuracy, particularly if we have a class distribution that is not uniform. It represents a weighted average of recall and precision. Due to this, both false positive and false negative results are considered. Accuracy is only the percentage of cases that were correctly predicted overall. Accuracy is a useful indicator of the performance of the model. Table 1 discusses the outcome of the proposed models in the testing phase.

Table I. Comparison of Proposed Models with Base Model

Model	Precision	Recall	F1-Score	Accuracy
CNN-Base Model	96.20%	77.55%	85.87%	84.17%
Model 1	93.67%	91.36%	92.50%	92.41%
Model 2	88.61%	89.74%	89.17%	89.24%
Model 3-SpinalNet	95.75%	97.82%	96.69%	97%

CNN base model resulted in the following performance metrics: precision of 96.20%, recall of 77.55%, F1-score of 85.87% and accuracy of 84.17%. Model 1 yielded the following performance metrics: precision of 93.67%, recall of 91.36%, F1-score of 92.50% and accuracy of 92.41%. Model 2 yielded the following performance metrics: precision of 88.61%, recall of 89.74%, F1-score of 89.17% and accuracy of 89.24%. SpinalNet generated the following performance metrics: precision of 95.75%, recall of 97.82%, F1-score of 96.69% and accuracy of 97%.

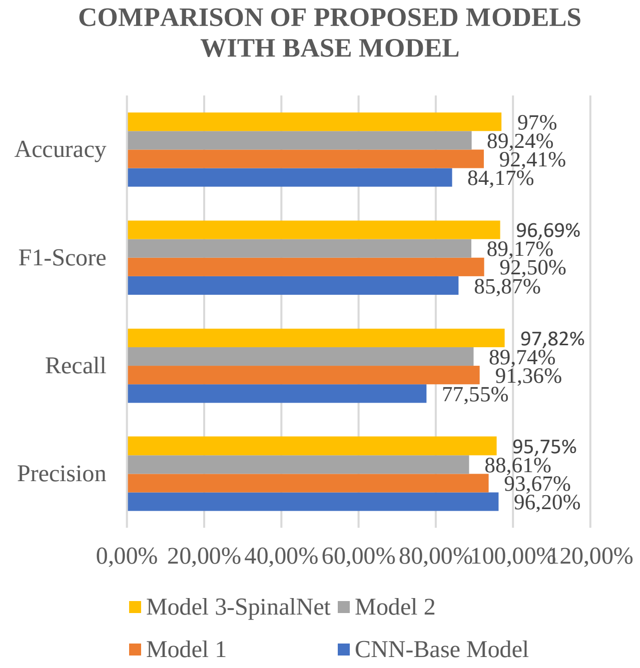

Figure 7. Comparison of proposed model with base model.

The recall, F1-Score, and accuracy of the three proposed models surpass those of cost of false positives is high, this statistic is essential since it increases the reliability of positive predictions. SpinalNet demonstrated superior performance in terms of recall, F1-Score, and the base CNN model. This indicates that, when compared to the base model, the presented models have been successful in enhancing the classification task's overall performance. On the other hand, the base CNN model exhibited the highest precision among the proposed models. This indicates that it is excellent at identifying positive instances. When the accuracy compared to the base CNN model. A high recall score means that a significant part of the real positive instances is successfully captured by SpinalNet. This is crucial in situations where it's imperative to find all positive cases, even if it means risking more false positives. Figure 7 shows comparison of proposed models with base models.

Among the three proposed models, SpinalNet consistently achieved the highest values across all four performance measures, showcasing its exceptional capabilities. High precision means that the system can recognize traffic accidents with accuracy, which is essential for avoiding erroneous warnings and unnecessary actions. By ensuring that accidents are almost certainly real when the system recognizes them, it minimizes false positives and subsequent disruptions. High recall enables the system to successfully record a significant number of actual traffic accidents. This is crucial since timely detection and response are crucial for emergency services and traffic management, and we want to ensure that the system doesn't miss any incidents. A high F1-Score indicates that the system achieves both accurate accident detection and good coverage of actual accident events. The system's overall correctness is essential to its dependability. The better accuracy of SpinalNet, which outperforms both the base CNN model and the other proposed models, implies that it produces accurate predictions over the whole dataset, which adds to its dependability and efficiency.

VIII. CONCLUSION AND FUTURE

A precise accident detection system is essential for improving emergency response and traffic safety. It enables speedy and effective emergency services, aids in better traffic management, increases driver awareness, and offers crucial information for enhancing road safety. The objective of this research is to propose deep learning models for detecting road accidents. Three efficient models were created, leveraging Machine Learning and Deep Learning techniques, namely VGG16, Inception V3, and SpinalNet. The use of SpinalNet represents a novel contribution. Notably, our experiments demonstrated an exceptional accuracy of 97% for SpinalNet, surpassing all other models evaluated in the study. This provides insights into the disparities in performance between SpinalNet and widely recognized pre-trained models.

We used an open dataset from Kaggle to develop the models. The dataset was collected from random YouTube videos that feature accidents and was divided into frames. After that, the frames were manually categorized as Accident or Non-Accident. So, we cannot completely rule out the possibility of biases that could have an impact on the validity of our findings because of the secondary data collection method.

Future studies can employ several datasets from different sources, encompassing diverse locations and timeframes. Additionally, incorporating real-time data from dependable traffic monitoring systems, authorized accident reports, and other trustworthy sources could also extend the dataset and boost the precision and dependability of the system used to detect road accidents. Also, model performance and convergence can frequently be improved by combining deep learning algorithms with optimization techniques. In the future, hybridizing SpinalNet with optimization techniques such as the Artificial Bee Colony (ABC) algorithm and Ant Colony Optimization (ACO) can offer promising opportunities for improving the performance of the road accident detection system.

A road accident detection system must achieve a balance between precision, recall, and accuracy to maximize road safety, improve emergency response times, minimize false alarms, avoid traffic jams, and pave the path for future improvements. By adopting cutting-edge technologies for accident detection, we can drastically lower the frequency of accidents, save lives, and make roads more secure and efficient for everyone.

References

[1] Road traffic injuries. [Online]. Available at: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries.

[2] W. Wijnen, "Socio-economic costs of road crashes in middle-income countries: Applying a hybrid approach to Kazakhstan," *IATSS Res.*, vol. 45, no. 3, pp. 293–302, 2021, https://doi.org/10.1016/j.iatssr.2020.12.006.

[3] Md. A. Fattah, S. R. Morshed, and A.-A. Kafy, "Insights into the socio-economic impacts of traffic congestion in the port and industrial areas of Chittagong city, Bangladesh," *Transp. Eng.*, vol. 9, p. 100122, 2022, https://doi.org/10.1016/j.treng.2022.100122.

[4] D. Oladimeji, K. Gupta, N. A. Kose, K. Gundogan, L. Ge, and F. Liang, "Smart transportation: An overview of technologies and applications," *Sensors*, vol. 23, no. 8, Art. no. 8, 2023, https://doi.org/10.3390/s23083880.

[5] Predictors of Mental Health Outcomes in Road Traffic Accident Survivors - PMC. [Online]. available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7074414/

[6] E. Poornima *et al.*, "Fog robotics-based intelligence transportation system using line-of-sight intelligent transportation," *Multimed. Tools Appl.*, 2023, https://doi.org/10.1007/s11042-023-15086-6.
[7] M. I. Basheer Ahmed *et al.*, "A real-time computer vision based approach to detection and classification of traffic incidents," *Big Data Cogn. Comput.*, vol. 7, no. 1, Art. no. 1, 2023, https://doi.org/10.3390/bdcc7010022.
[8] H. Hozhabr Pour *et al.*, "A machine learning framework for automated accident detection based on multimodal sensors in cars," *Sensors*, vol. 22, no. 10, Art. no. 10, 2022, https://doi.org/10.3390/s22103634.
[9] N. Pathik, R. K. Gupta, Y. Sahu, A. Sharma, M. Masud, and M. Baz, "AI enabled accident detection and alert system using IoT and deep learning for smart cities," *Sustainability*, vol. 14, no. 13, Art. no. 13, 2022, https://doi.org/10.3390/su14137701.
[10] I. Badi, M. B. Bouraima, and M. Jibril, "The role of intelligent transportation systems in solving traffic problems and reducing environmental negative impact of urban transport," p. 2023, 2022, https://doi.org/10.55976/dma.1202311371-9.
[11] M. M. Rahman, M. K. Islam, A. Al-Shayeb, and M. Arifuzzaman, "Towards sustainable road safety in Saudi Arabia: exploring traffic accident causes associated with driving behavior using a Bayesian belief network," *Sustainability*, vol. 14, no. 10, Art. no. 10, 2022, https://doi.org/10.3390/su14106315.
[12] V. Petraki, A. Ziakopoulos, and G. Yannis, "Combined impact of road and traffic characteristic on driver behavior using smartphone sensor data," *Accid. Anal. Prev.*, vol. 144, p. 105657, 2020, https://doi.org/10.1016/j.aap.2020.105657.
[13] S. Sahu, S. Mishra, K. K. Barik, and D. Sahu, "Implementation of road safety audit to highlight the deformities in the design and environmental safety features: A case study on national highway-326," *Int. J. Environ. Clim. Change*, vol. 12, pp. 1123–1140, 2022, https://doi.org/10.9734/ijecc/2022/v12i1131089.
[14] J. El, "Historical developments of random forest," *Medium*, Jul. 23, 2020. [Online]. available at: https://drjariel.medium.com/historical-developments-of-random-forest-41492deb6737.
[15] explorium_admin, "Decision Trees: Complete Guide to Decision Tree Analysis," *Explorium*, Dec. 10, 2019. [Online]. Available at: https://www.explorium.ai/blog/the-complete-guide-to-decision-trees/
[16] K. Thankachan, "What? When? How?: ExtraTrees Classifier," *Medium*, Aug. 09, 2022. [Online]. Available at: https://towardsdatascience.com/what-when-how-extratrees-classifier-c939f905851c.
[17] J. S. Cramer, The origins of logistic regression, 2002. https://doi.org/10.2139/ssrn.360300.
[18] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," *Proceedings of the 2017 International Conference on Engineering and Technology (ICET)*, August 2017, pp. 1–6. https://doi.org/10.1109/ICEngTechnol.2017.8308186.
[19] J. G. Choi, C. W. Kong, G. Kim, and S. Lim, "Car crash detection using ensemble deep learning and multimodal data from dashboard cameras," *Expert Syst. with Appl.*, vol. 183, p. 115400, 2021, https://doi.org/10.1016/j.eswa.2021.115400.
[20] N. R. Vatti, P. L. Vatti, R. Vatti, and C. Garde, "Smart road accident detection and communication system," *Proceedings of the 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, March 2018, pp. 1–4. https://doi.org/10.1109/ICCTCT.2018.8551179.
[21] H. M. Sherif, M. A. Shedid, and S. A. Senbel, "Real time traffic accident detection system using wireless sensor network," *Proceedings of the 2014 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR)*, August 2014, pp. 59–64. https://doi.org/10.1109/SOCPAR.2014.7007982.
[22] J. Amala Ruby Florence and G. Kirubasri, "Accident detection system using deep learning," in *Computational Intelligence in Data Science*, L. Kalinathan, P. R., M. Kanmani, and M. S., Eds., in IFIP Advances in Information and Communication Technology. Cham: Springer International Publishing, 2022, pp. 301–310. https://doi.org/10.1007/978-3-031-16364-7_23.
[23] S. Robles-Serrano, G. Sanchez-Torres, and J. Branch-Bedoya, "Automatic detection of traffic accidents from video using deep learning techniques," *Computers*, vol. 10, no. 11, Art. no. 11, 2021, https://doi.org/10.3390/computers10110148.
[24] A. K. Paul, P. K. Boni, and M. Z. Islam, "A data-driven study to investigate the causes of severity of road accidents," *Proceedings of the 2022 IEEE 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2022, pp. 1–7. https://doi.org/10.1109/ICCCNT54827.2022.9984499.
[25] T. K. Vijay, D. P. Dogra, H. Choi, G. Nam, and I.-J. Kim, "Detection of road accidents using synthetically generated multi-perspective accident videos," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 2, pp. 1926–1935, 2023, https://doi.org/10.1109/TITS.2022.3222769.
[26] K. V. Thakare, D. P. Dogra, H. Choi, H. Kim, and I.-J. Kim, "Object interaction-based localization and description of road accident events using deep learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 11, pp. 20601–20613, 2022, https://doi.org/10.1109/TITS.2022.3170648.
[27] A. Saveliev, V. Lebedeva, I. Lebedev, and M. Uzdiaev, "An approach to the automatic construction of a road accident scheme using UAV and deep learning methods," *Sensors*, vol. 22, no. 13, Art. no. 13, 2022, https://doi.org/10.3390/s22134728.
[28] M. S. Basit, U. Ahmad, J. Ahmad, K. Ijaz, and S. F. Ali, "Driver drowsiness detection with region-of-interest selection based spatio-temporal Deep Convolutional-LSTM," *Proceedings of the 2022 16th International Conference on Open Source Systems and Technologies (ICOSST)*, 2022, pp. 1–6. https://doi.org/10.1109/ICOSST57195.2022.10016825.
[29] P. Prajwal, D. Prajwal, D. H. Harish, R. Gajanana, B. S. Jayasri, and S. Lokesh, "Object detection in self driving cars using deep learning," *Proceedings of the 2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, September 2021, pp. 1–7. https://doi.org/10.1109/ICSES52305.2021.9633965.
[30] T. Tamagusko, M. G. Correia, M. A. Huynh, and A. Ferreira, "Deep learning applied to road accident detection with transfer learning and synthetic images," *Transp. Res. Procedia*, vol. 64, pp. 90–97, 2022, https://doi.org/10.1016/j.trpro.2022.09.012.
[31] Accident Detection from CCTV Footage. [Online]. Available at: https://www.kaggle.com/datasets/ckay16/accident-detection-from-cctv-footage.
[32] R. Chauhan, K. K. Ghanshala, and R. C. Joshi, "Convolutional Neural Network (CNN) for image detection and recognition," *Proceedings of the 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, December 2018, pp. 278–282. https://doi.org/10.1109/ICSCCC.2018.8703316.
[33] S. Indolia, A. K. Goswami, S. P. Mishra, and P. Asopa, "Conceptual understanding of convolutional neural network - A deep learning approach," *Procedia Comput. Sci.*, vol. 132, pp. 679–688, 2018, https://doi.org/10.1016/j.procs.2018.05.069.
[34] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010, https://doi.org/10.1109/TKDE.2009.191.
[35] S. Tammina, "Transfer learning using VGG-16 with deep convolutional neural network for classifying images," *Int. J. Sci. Res. Publ. IJSRP*, vol. 9, no. 10, p9420, 2019, https://doi.org/10.29322/IJSRP.9.10.2019.p9420.
[36] X. Xia, C. Xu, and B. Nan, "Inception-v3 for flower classification," *Proceedings of the 2017 2nd Int. Conf. Image Vis. Comput. ICIVC 2017*, pp. 783–787, 2017, https://doi.org/10.1109/ICIVC.2017.7984661.
[37] J. Tao, Y. Gu, J. Sun, Y. Bie, and H. Wang, "Research on vgg16 convolutional neural network feature classification algorithm based on Transfer Learning," *Proceedings of the 2021 2nd China International SAR Symposium (CISS)*, 2021, pp. 1–3. https://doi.org/10.23919/CISS51089.2021.9652277.
[38] A. Bagaskara and M. Suryanegara, "Evaluation of VGG-16 and VGG-19 deep learning architecture for classifying dementia people," *Proceedings of the 2021 4th International Conference of Computer and Informatics Engineering (IC2IE)*, September 2021, pp. 1–4. https://doi.org/10.1109/IC2IE53219.2021.9649132.
[39] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826. https://doi.org/10.1109/CVPR.2016.308
[40] C. Wang *et al.*, "Pulmonary image classification based on inception-v3 transfer learning model," *IEEE Access*, vol. 7, pp. 146533–146541, 2019, https://doi.org/10.1109/ACCESS.2019.2946000.
[41] H. M. D. Kabir *et al.*, *SpinalNet: Deep Neural Network with Gradual Input*. 2020.

SHERIMON P.C. received his M.C.A. from Madras University and Ph.D. in Computer Applications from Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu in 2016. Currently he is the dean of Faculty of Computer Studies at Arab Open University, Muscat, Sultanate of Oman.

His research interests lie in semantic web, ontology, machine learning, and deep learning.

VINU SHERIMON received her M.Sc. in Computer Science from School of Computer Sciences, Mahatma Gandhi University, Kerala and Ph.D. in Computer Science from Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu in 2016. Currently, she is working at College of Computing and Information Sciences, University of Technology and Applied Sciences, Muscat, Sultanate of Oman.

Her research interests lie in semantic web, ontology, data mining, machine learning, and blockchain.

JOHNSYMOL JOY holds an M.Tech in Computer Science and Technology and is currently a Ph.D. candidate in the Computer Science and Engineering Department. Her research interests include processing and analyzing different neuroimaging biomarkers using Deep Learning techniques for detecting various neurodevelopmental disorders. She works as an Assistant Professor in the PG Department of Computer Applications and Artificial Intelligence at Saintgits College of Applied Sciences, Kerala.

AMBILY MERLIN KURUVILLA is an associate Professor at Computer Applications and Artificial Intelligence department of Saintgits College of Applied Sciences, Kottayam. She holds MCA, M.Phil and Ph.D. degrees in Computer Science. She has more than 19 years of teaching experience and has published research papers in Journals and Proceedings of International repute.

ARUNDAS G. holds an M.Sc. in Artificial Intelligence from Saingits College of Applied Sciences (2021-2023) and BCA degree from Sree Narayana Guru Institute of Science and Technology (2018-2021). He is a Magento developer at Ceymox, contributing to e-commerce development. His research interests include machine learning and image processing.