Street-crimes modelled arms recognition technique employing deep learning and quantum deep learning (SMARTED)

Syed Atif Ali Shah[1,2], Ahmed Abdel-Wahab[3,4], Nasir Ageelani[2], Najeeb[2]
[1]Department of Computer Science, Faculty of Computer Science and Artificial Intelligence, Air University, Islamabad, Pakistan
[2]Department of Computer Science, Faculty of Computer and Information Technology, Al-Madinah International University, Kuala Lumpur, Malaysia
[3]Department of Computers and Systems Engineering, Faculty of Engineering, Al-Azhar University, Cairo, Egypt
[4]Faculty of Computer Studies, Arab Open University, Riyadh, Saudi Arabia

Article Info

ABSTRACT

An increase in population causes loopholes in controlling law and order situations. One of the threatening aspects of peace is the availability of weapons to the general public. As a result, many dangerous situations arise, most notably street crimes. Traditional methods are not sufficient to deal with such situations. Consequently, the police and other security concerns need serious technological reforms to prevent such situations. In modern technology tools, deep learning has made great improvements in various areas of daily life, especially in object detection. This paper presents an efficient technique for detecting weapons from closed-circuit television (CCTV) cameras, videos, or images. Upon the detection of the weapon, the concerned person is automatically informed to take the necessary action; without human intervention. For the first time, RetinaNet has been employed to detect weapons in real-time scenarios. RetinaNet has shown remarkable improvement in this domain, by achieving an average of 90% accuracy in real-time scenarios. With the emergence of quantum computing, many computer environments saw a revolution. Thus, we have also utilized quantum computing technology for real-time weapons detection using quantum deep learning. In this paper, quantum inspired RetinaNet (QIR-Net) is developed for weapons detection and amazing results are observed.

Corresponding Author:

Syed Atif Ali Shah
Department of Computer Science, Faculty of Computer Science and Artificial Intelligence, Air University
Islamabad, Pakistan
Email: atif.ali@mail.au.edu.pk

1. INTRODUCTION

In current circumstances, there is no single source of true data on violence in the world. Besides the Geneva declaration secretariat, the United Nations office on drugs and crime (UNODC) and the world health organization (WHO) are persistently collecting information on violence, which takes place due to different reasons; to identify the patterns in armed violence. Typically, the impact of armed violence is estimated through the number of deaths that occurred due to violence. Figure 1 shows the type of violent death as defined by the Geneva declaration (2015 edition). In third-world countries, the data collected are far from the facts. The whole cry is because of the availability of weapons for the public. Consequently, many dangerous situations have emerged, the most common of which is street crime. Traditional methods are not sufficient to address such situations. Police and other institutions, therefore, use advanced technology to maintain social peace.
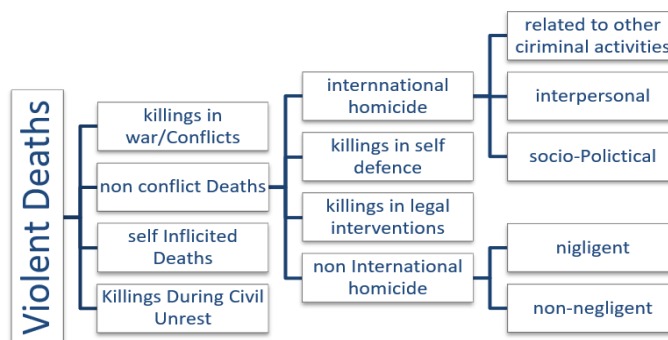
Figure 1. Violent deaths by the geneva declaration

Modern technology focuses on the detection of pistols. None of the studies addressed the question of detecting several types of weapons, i.e. commonly used in street crimes. In addition, the work featured in the existing literature uses the kind of armory used in Hollywood. The problem arises when the same model is applied in a different context. For example, the handgun used in third-world countries is different in shape than in Hollywood. This is why the model will be less accurate in detecting when it is implemented in other parts of the world. There is also a need for improvement in terms of accuracy and speed. However, the proposed models will detect all types of armory commonly used in street crimes; worldwide. This paper will also investigate a novel convolutional neural network (CNN) model, i.e. RetinaNet, which has higher accuracy and higher speed than existing state-of-the-art object detection CNN models. To overwhelm the existing issues; a quantum deep learning model is being developed that can detect and localize armory in images/videos in real-time. Street-crimes modelled arms recognition technique employing deep learning and quantum deep learning (SMARTED) can be used in many ways to overcome law and order situations and could be an astonishing revolution for instilling peace in society. For example, it can be used for public safety in public safety i.e offices, govt offices, courts, schools, institutions, universities, mosques, churches (worship and sacred places), bus stands, airports, parks, playgrounds, stadiums, sports complexes, shopping malls, markets, hotels, restaurants, food streets, hospitals, clinics, and especially in sensitive places.

It is not as expensive as X-ray technology nor as slow as other automatic machines to only work on images. It can be used with a closed-circuit television (CCTV) camera to detect armories in real-time videos. After detecting such weapons, it automatically alerts police and other security agencies to take the necessary action without spending much time. As a result, the rate of violent and criminal activity in the general public has declined dramatically. The contribution of this paper is:

- This study is the first to apply the convolutional neural networks and quantum inspired convolutional neural networks i.e RetinaNet and quantum inspired RetinaNet (QIR-Net) models to detect weapons from videos and images.
- An improved technique for the detection of weapons is being studied, with an accuracy of about 90%; this has never been seen before.
- These models can detect all types of weapons used in street crimes worldwide. Unlike existing techniques, those were only capable of detecting pistols or knives and fewer of detecting handguns.
- A new data set is utilized that contains all types of weapons used in street crime.
- Research is not limited to images but mainly focuses on videos and scenes in real-time.
- Implementation is not limited to specific areas but can be applied anywhere world.

This paper has been divided into six sections. The first section describes the basic challenges and problems associated with the existing solutions. The second section summarizes the historical background. The third section describes the technical details of the proposed solution, which contains both solutions, i.e., conventional computing and quantum computing technologies. It also describes the proposed architecture and operations of the models. The fourth section presents the results. Section five demonstrates the entire research and future work, finally sixth section is acknowledgement.

2.   LITERATURE REVIEW

The journey began with a manual system and then reached a fully automated [1] and intelligent system [2]. In this paper, the study started from the manual approach which includes the weighted mixture of Gaussians [3], polarized signals-based technique, a multiresolution mosaic technique, and three-dimensional (3D) computed tomography (CT) [4] Haar cascades procedure. Advancements began with machine learning methodologies

such as the visual background extractor algorithm, three-layered artificial neural network (ANN) [5] employed with active millimeter waver radar, X-ray-based technique [6], and ANN-based MPEG-7 classifier.

Nowadays, Santaquiteria *et al.* [7] preventing security risks is critical, and deep learning-based technologies enable the development of automatic weapon detectors. Muchiri *et al.* [8] research; the authors presented the approach to combining information on weapon appearance and human position [9]. The important parts of the pose are to extract the hand region and then process it in separate subnetworks in order to merge the handgun. The score ranged from 4.23 to 18.9 points on the AP scale. Hernández *et al.* [10] to an increase in crime, security is always a key concern in all domains, and surveillance systems are used for a variety of purposes. For automatic weapon detection; the researchers use two types of [11] algorithms (CNN and region-based CNN (RCNN)) [12], [13]. Photos that have been pre-labeled as well as images that have been manually labeled are used. The resulting algorithms are accurate, but in practice, there is a trade-off between [14] speed and accuracy. In [15] the researcher develops a fully automated system to recognize basic weaponry, such as handguns and rifles and applies it to our surveillance system utilizing you only look once (YOLO) V2 and V3 models [16] to detect a weapon or unsafe assets to avoid any type of assault or risk to human life. Kanehisa and Neto [17] and Uddin *et al.* [18] artificial intelligence (AI) [19], computer vision is used to recognize patterns.

Hao *et al.* [20] research entails a comparison of various online handgun detection technologies. Supervised deep learning is used in this method. Elmir *et al.* [21], [22] the authors presented several surveys that revealed that the handgun is the most widely used weapon for a variety of crimes such as burglary, rape, and so on. In the current scenario, automatic gun recognition from a congested scene utilizing convolutional neural networks (CNN) [23] is too important. Shah *et al.* [24] automatic gun recognition from congested scenes, they employed a deep-convolutional network with transfer learning. Concerning the Verma and Dhillon [25] internet movie firearms data base (IMFDB), we have seen an excellent performance from the system. Weapons [26] classification and anomalous event detection can be done using CCTV systems. The suggested project uses three processing modules: one for object detection, another for weapon categorization, and a third for alert, monitoring, and control. Weapons classification [27] and anomalous event detection can both be done using CCTV systems. Lu [28] suggested by the researchers using a state-of-the-art model called YOLO V3 and YOLO V4 will automatically monitor and perform all operations in a circular region, reducing crime and providing a higher level of protection. Pang *et al.* [29] project uses three processing modules: one for object detection, another for weapon categorization, and a third for alert, monitoring, and control. Lamas *et al.* [30] research; authors presented a new algorithm that detects firearms in videos automatically [31] and produces acceptable results in low-quality videos. The authors also introduce the alarm activation time per interval (AATpI), a new metric for evaluating the effectiveness of a detection model as an automatic video detection system.

Tiwari *et al.* [32] research; the authors presented convolutional neural networks (CNN) and a brightness-directed preprocessing approach that have improved the automatic recognition of cold steel weapons handled by one or many people in video surveillance darkening and contrast at learning and test stages (DaCoLT). Object [33] detection models have improved in recent years, yet in low-quality videos, they still create false positives. To reduce the number of false positives in surveillance footage, this research introduces a novel picture fusion approach. Piyadasa [34]; the researchers suggested employing a two-level deep learning methodology to improve the recognition of small objects that are handled identically. They concentrate on detecting weapons and things that could be mistaken for a handgun or a knife in video surveillance. Gelana and Yadav [35] researchers' use of single-image detectors in real-time firearm alarm systems; is still questionable, despite developments in computer vision. Multicast, a new approach based on convolutional neural networks and long-short-term memory (LSTM) networks [9], can cut false alarms by 80%. Danilov *et al.* [36] traditional CCTV surveillance and control system necessitates human resource supervision, and automatic identification of weapons is a critical requirement for surveillance development. In this research; the authors focused on the implementation and comparison of quicker techniques with visual geometry group (VGG) and residual neural network (ResNet), such as region (R-CNN) and region fully convolutional networks (R-FCN). González *et al.* [37] developing countries, the rate of street crime is extremely high. CCTV cameras are sometimes used to keep an eye on visual armory, but in this research; the authors focus on deep learning using CNN's architectures that revolutionized the weapon identification system. Cold steel [32] weapons can be detected automatically and prevent crime, but their surface reflection makes their outlines in surveillance videos unclear. In this study, the authors have explained that this can be avoided by using a light-oriented preprocessing approach. Due to the unavailability [38] of quantum machines/devices, there is a huge gap in development and research.

Mattern *et al.* [39] proposed a technique to implement the quantum-based model on conventional computing models. Here K-means clustering algorithms are used to assure the working. Taha and Nawar [40] machines are not fast enough to meet the actual requirements, the results shown results are better than

conventional programming. NGSA-ii [41] based quantum support vector machine is used on multi-objective genetic algorithms. Hur *et al.* [42] the authors used a non-linear dataset and found better results. Lockwood and Si [43] to increase the performance of classical CCN models, the authors proposed updating the general functions with the potential of parametrized quantum circuits (PQCs). Hassan *et al*. [44] proposed model is trained and tested on the modified national institute of standards and technology (MINIST) dataset and higher accuracy is found as a result. Zhou *et al*. [45] Introduce a deep quantum neural network by adding quantum-structured layers to the deep neural network. Because of continuity, it represents more features and benefits in non-linearity.

3. METHODOLOGY

This section introduces the tools and techniques used to design and implement the proposed model. This study includes the making dataset proceeded by development of and training of model, as the training of the model is repetitive process, and finally the result evaluations. A complete roadmap is described to show the workings of the proposed model.

3.1. Dataset

Deep learning models require large amounts of data to train the models. There are mainly two challenges when selecting the datasets. First, existing datasets do not cover the range of weapons used in street crimes worldwide. Second, the existing datasets contain the weapons that are generally seen in films, but in real situations, it is completely different. Consequently, even more, performant models fail in real-time cases. With all this in mind, a newly developed dataset, the street crime arms data set (SCAD) is utilized. SCAD contains four types of arms.

- Short-range rifle Ak47.
- Shotgun.
- Pistol.
- Knife.

Approximately 4,000 images are collected and distributed equally to all types. Of these images, between 70% and 80% are used for the training model and the rest for testing. Sample images are shown in Figure 2. Data collected from security departments like police, law enforcement agencies, and security agencies.

Figure 2. Sample weapons from SCAD

3.2. Proposed methodology

A detailed model is presented here. Initially, the data sets are designed and processed to make them workable for the training of the model. The Figure 3 summarize the proposed method. Then the RetinaNet model architecture is described in detail. How to get data from the input process and how to predict. This discussion covers the performance and features of the RetinaNet. To show the results of RetinaNet, different performance measurement tools are used to understand the functionality of the model. Some of the data were also tested outside the data set to validate the overall performance of the RetinaNet. The conclusions describe the use of RetinaNet for weapon detection and how they can be used in other areas for future prospects.

3.3. The retinanet

Before the RetinaNet, there was a long debate about the performance of detectors in single-stage and two-stage. One-stage detectors were thought to be more efficient than two-stage detectors, but no such examples can be cited; due to the high background and foreground imbalances of the detector of one stage

detector. As a single-stage detector, it outperformed the models in the same class. RetinaNet is divided into two modules, the first module is a backbone called feature extractor. The second is made up of two subnetworks known as predictor. The first module, the feature extractor, is further divided into bottom-up and top-down paths. The bottom-up path consists of ResNet, a down-sampling process, and FPN, while the top-down path consists of an up-sampling process and the instigation of features. However, the second module, predictor or classification Subnet, is made up of bounding box regression and classification subnet. Figure 4 shows the basic architecture of RetinaNet, which consists of three layers. The first layer called the bottom-up path includes ResNet as the backbone and process of down-sampling at each level, this down-sampling at each layer is collectively called feature pyramid network. The second layer is top down path, which is further divided into up-sampling and proceeded by feature integration. Finally, the third layer called the predictor's head is composed of the bounding box and classification.
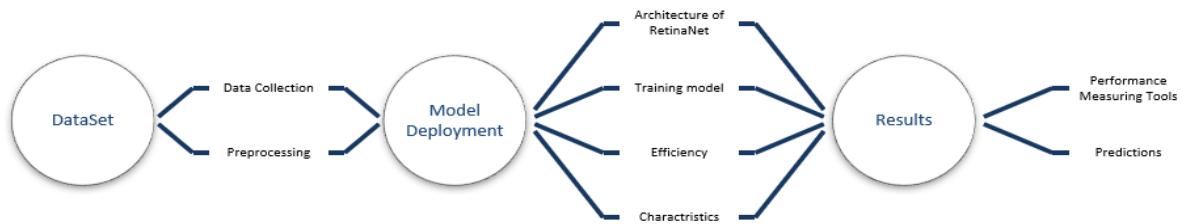
Figure 3. Proposed methodology
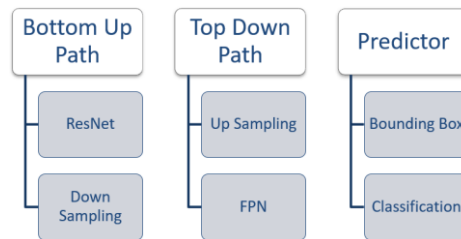
Figure 4. Basic architecture of RetinaNet

3.3.1. Bottom-up path
The image is first put into ResNet 101, which uses the 1st convolution module with stride 2 to cut the size of an image in a half, with a filter/kernel size of 64 and stride 2. Through this process the though the size of the image decreases but it provides enrich feature map. However, it does not lower the overall speed of object detection process. Details of the procedure are described below.

3.3.2. Top-down path
After the bottom-up phase has been completed, the top-down process begins, as depicted in Figure 5. The characteristics of C5 are transmitted to M5, and the actual up-sampling process begins here. To get an enhanced feature map, features from the previous layers are added element-by-element (combined), then features from M5 are doubled in size and merged with 1x1 convolution of C4, and then transferred to M4, a technique known as up-sampling. The feature map's size is doubled for each M(x) to make it compatible with each, 1x1 convolution of C(x). Moving from M4 to M3 and then to M2 follows the same approach.
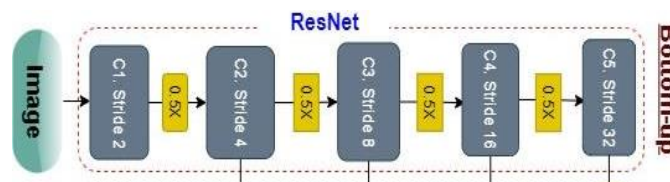

Figure 5. Down-sampling

3.3.3. Up-sampling

The retina net employs an up-sampling approach to overcome this constraint. The features of the previous layer are integrated into this procedure, resulting in an enriched feature map. The up-sampling technique is established in two stages, with the bottom-up and top-down paths being followed. Down-sampling is used to extract desired features in the bottom-up path, and then up-sampling is used to increase the density of desired features in the top-down path.

3.3.4. Feature extractor

Finally, all fused layers are subjected to three convolutions. When combined with the up-sampled layer, this filter eliminates aliasing. As discussed, 3×3 convolution is applied to M4, and P4 is generated as a result. The feature pyramid network (FPN) is a feature extractor that operates at high speed while maintaining high accuracy.

3.3.5. Feature pyramid network

FPN or feature pyramid network is an essential part of this model. The process of feature extraction at each level is performed in such a way that it seems like a pyramid. Features from the first module are extracted, then reduced to half, and passed to the second module. The same procedure is applied to each module thus feature map is reduced half by each module. This reduction in the feature maps results in the finding of only important features. Thus, the resultant is enriched feature map. This feature pyramid network is achieved using down-sampling and up-sampling or bottom-up and top-down approach.

3.3.6. Task-specific networks

We have two networks in the RetinaNet model: a backbone network and two sub-networks called task-specific networks. The image's feature map is computed by the backbone. The backbone can be built using traditional CNN designs like ResNet50 or ResNet101. In the backbone, RPN (region proposal network) [46], feature pyramids [37], and FPN [7] are used (as discussed above).

3.3.7. Classification and bounding box

The first subnet does the categorization and forecasts the likelihood of an object's presence. Bounding boxes are computed in the second subnet. The output is then subjected to a sigmoid activation function and focal loss. Figure 6 depicts the RetinaNet model.
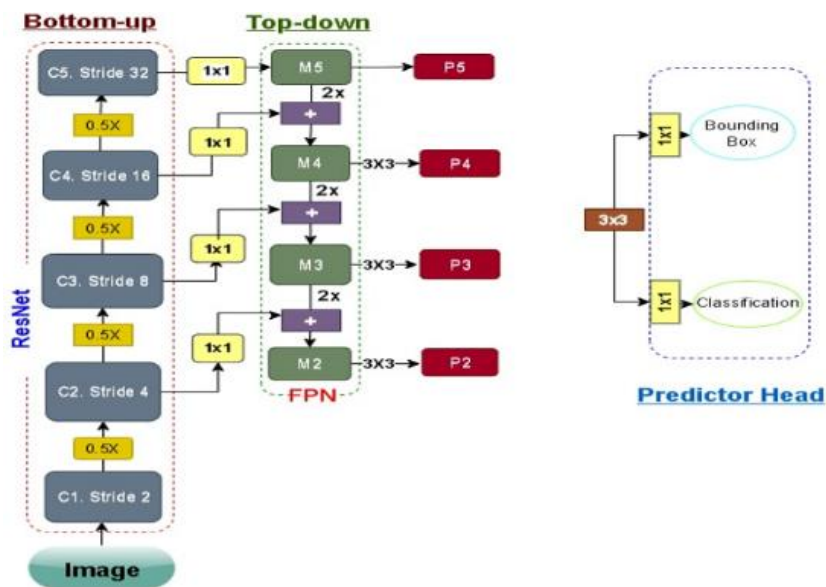

Figure 6. The model of RetinaNet

3.4. Quantum inspired convolutional neural network

The quantum inspired convolutional neural networks are intermediaries between pure Qonvolutional (also know as Quantum Convolutional) neural networks and convolutional neural networks. As we all know about the availability of quantum computers, there are few. A sort of personal quantum computer is not publicly

available yet. That's why it is not easy to program and test quantum deep learning models. Therefore, quantum inspired convolutional neural networks (QICNNs) are designed to rescue us and help us to develop quantum deep learning models.

3.4. Model description

A detailed model is presented here. Initially, the data sets are designed and processed to make them workable for the training of the model. Figure 7 summarizes the functionality of the model. Then the QIR-Net model architecture is described in detail. How to obtain data from the input process and predict it. This discussion deals with QIR-Net performance and features. In order to ensure that the functionality of the model is implemented, different performance measurement tools are used to understand the functionality of the model. Some data were also tested outside the data set to verify the overall performance of QIR-Net. The conclusions describe how the quantum computing model can be used for weapon detection and how it can be used for future prospects in other areas.
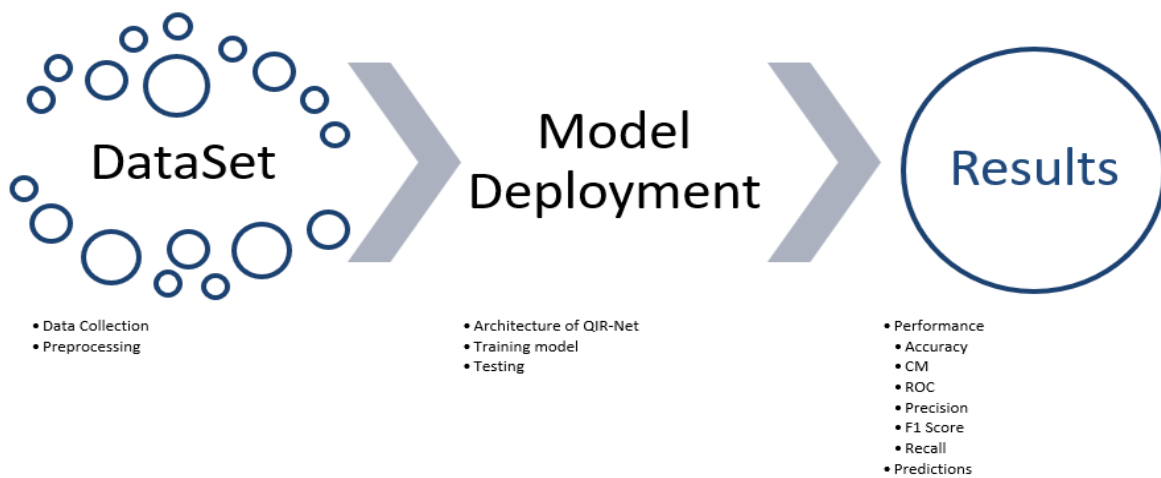
Figure 7. Summary of model

3.5. The QIR-Net

This model is a stepping stone in our research to put our tools and technologies in the modern territory of quantum computing. As discussed earlier we won't be able to truly put our work into a quantum domain, though it will take time. However, we made an effort to take a step in that direction. To get into quantum computing we proposed a model that combines the working of quantum computing with deep learning (i.e RetinaNet), thus named it QIR-Net the quantum inspired RetinaNet. QIR-Net is divided into two modules, the first module is a backbone called feature extractor. The second is made up of two subnetworks known as predictor. As we discussed earlier in RetinaNet, our inspiration is RetinaNet. That's why we tried to make a similar model, i.e., having the same architecture even after introducing the quantum computing approach. The first module is modified and in this module, the quantum convolutional neural network is implemented. This 1st module is practically quantum.

3.5.1. Feature extractor

The feature extractor is the first module of QIR-Net, which is mainly responsible for extracting features from the input image. This is not a simple process, as it is composed of many sub-processes. In any CNN model, feature extraction is the most complex and core task in the whole model. in other words, the accuracy and speed of the model depend on it. The more features we may have, the better the accuracy level will be expected, however, at the same time higher computational complexity may also arise. To reduce computational complexity and increase accuracy, we only selected this module to be handled using quantum computing machinery. Therefore, our proposed model will show higher accuracy along with a greater speed. To extract the desired features, we may pass through different processes as described in Figure 8.
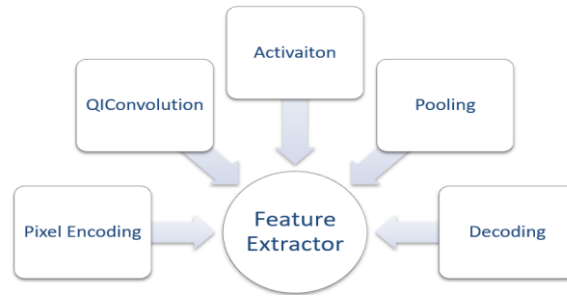
Figure 8. Feature extractor

3.5.2. Encoding

In quantum computation, the quantum feature map is used to convert input data to a Hilbert space; φ: X → H. Hilbert spaces are vector spaces that represent the state of quantum mechanics' physical systems. This is particularly important for quantum computing because our data is now represented by vectors in the Hilbert space. For such conversion, we have four techniques.

- Qbit encoding.
- Amplitude encoding.
- Dense encoding.
- Hybrid encoding.

Usually, amplitude encoding is used because of its simplicity and speed, however, qbit is more natural to use in many cases. Amplitude encoding means that input data are encoded as probability amplitudes of a quantum state. When encoding the input data amplitude, the number of training parameters of the QNN O(logn) scale is appreciably lower than O(n), so amplitude encoding is so powerful.

3.5.3. Quantum-inspired convolutions

QICNNs use qbits to process data on conventional computers while acting like real QCN models. In this technique, quantum-inspired neurons are utilized to develop a model. See Figure 9 for a model of quantum inspired neuron (QIN) [38]. First part i.e Figure 9(a) shows the Structures of common quantum-inspired neurons, on this basis, a typical model of the neuron inspired by quantum can be obtained. In previous work, the weight (u_i) is considered to be a real number because it is convenient to implement it in the common framework of machine learning. However second part Figure 9(b) shows the improved quantum-inspired neurons with real value. The complex value networks have a richer representation capacity and a better nonlinearity. Therefore, it can directly use complex weights to improve the neurons inspired by quantum.

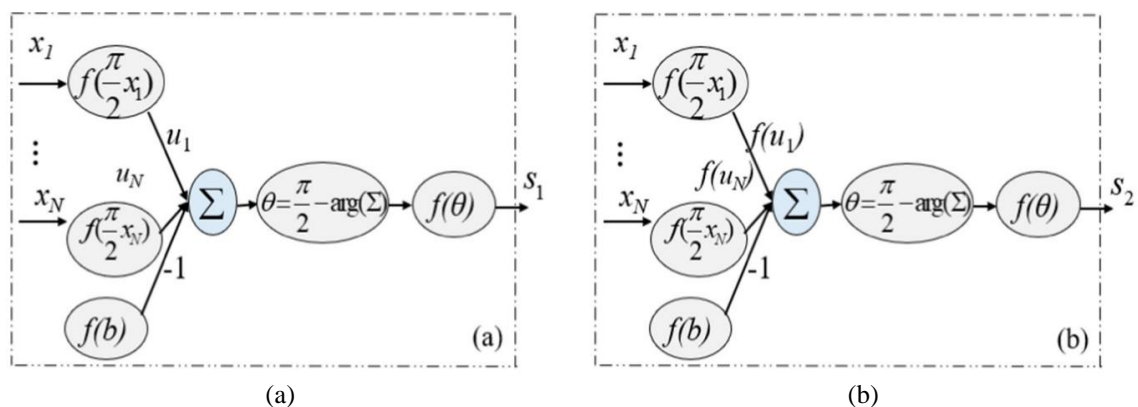(a)   (b)

Figure 9. Structure of the common quantum-inspired neuron (a) structures of common quantum-inspired neurons and (b) improved quantum-inspired neurons with real value

The weights {u1, u2…uN} are real numbers. The module Σ represents the weighted summation operation. The quantum-inspired method of implementing the convolutional operations is used to construct quantum-inspired convolutional neural networks. It is worth mentioning that, though the implementation is

somewhat more complicated than the conventional CNN model, results are improved. The easy way to build QICNNs is to use quantum-inspired neurons in fully connected layers and use quantum-inspired convolutional operations in convolutional layers. Each QIC quantum-inspired convolution layer is followed by an activation function; here rectified linear units (ReLU) is used; And then the pooling layer. This group of layers i.e convolution, ReLU, and pooling layer. It was repeated multiple times, as much as we are interested in extracting features. But higher depth increases complexity more than expected features, and thus faster machines are also required. To reduce the complexity and extract sufficient features to make detection feasible, we repeat it three times, as shown in Figure 10. Finally, we'll get the strong feature map, but this feature map is in qbits. Therefore, we need to convert them back to conventional bits. This process is called decoding and is the last step of a quantum-inspired convolutional neural network.
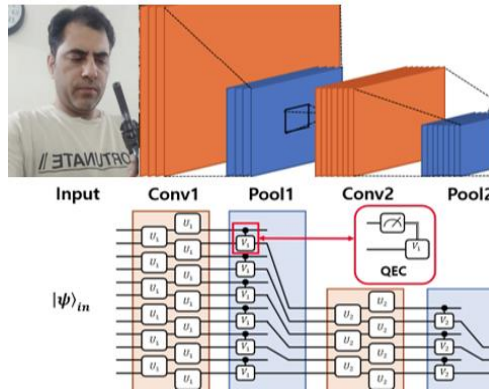
Figure 10. Summary of the QIR-Net feature extraction process

3.5.4. Predictor head
Now at this stage, the information is the same as we got in RetinaNet from the feature extraction module. Thus, the rest of the process is the same as we discussed earlier. Similarly, here also we have a backbone network and two subnetworks called task-specific networks. The image's feature map is computed by the QICNN.

3.5.5. Classification and bounding box
The first subnet does the categorization and forecasts the likelihood of an object's presence. Bounding boxes are computed in the second subnet. The output is then subjected to a sigmoid activation function and focal loss. The process model of QIR-Net is described in Figure 11, this diagram shows the detailed working of the model. how features are extracted from the image using quantum computing and the using predictor head results are presented.
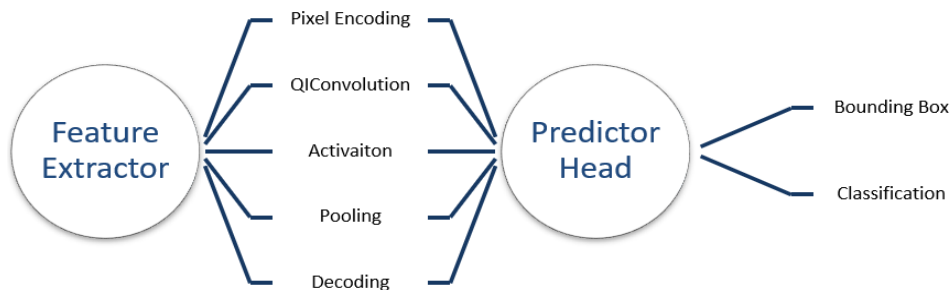

Figure 11. The QIR-Net model

4. RESULTS
Finally, in this section, the performance and accuracy of the model are presented. To validate the working of the proposed model, different evaluation techniques are employed. These evaluation techniques are

discussed below. The model of a convolutional neural network is trained on training data. Validation data is used to tune and retrain the model to optimize the output.

4.1. Accuracy

Accuracy shows the performance of architecture among all the classes; however, it is more reasonable when multiple classes are having the same significance, as shown in Figure 12. In terms of formula, it is the ratio between accurate results and total outcomes. It shows the possible outcomes and, hence, shows the performance of RetinaNet and QIR-Net compared to other state-of-the-art models. Accuracy is not the only feature that has been calculated for showing the performance of both models i.e RetinaNet and QIR-Net but there are other features to present the outperformance of these models, discussed ahead. But among all those factors, accuracy is the most commonly used term and is also easy to understand at first glance.
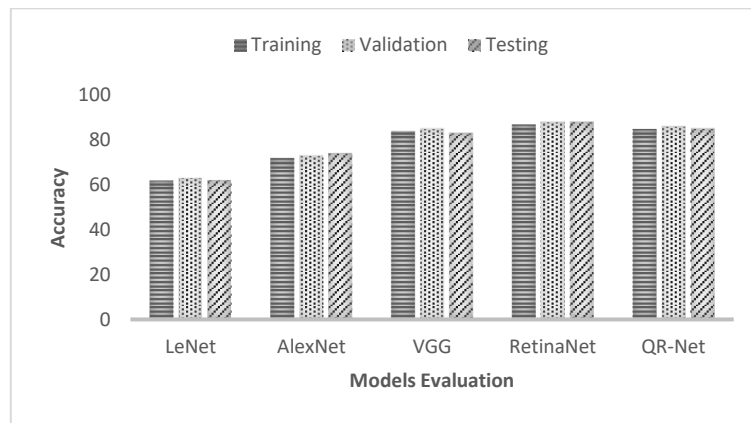
Figure 12. Accuracy in % of five models

4.2. Confusion matrix

A confusion matrix is an alternative method to demonstrate the performance of the classification model, used. Higher diagonal values mean higher accuracy, as presented in Figures 13 to 17 for each model respectively. The confusion matrix is an important and useful tool in statistics and information handling, especially when illustrating Type I and Type II errors. Through these errors (factors), the reliability of the system/model/outcomes can be directly addressed from the confusion matrices of five models ie, LeNet, AlexNet, VGG, RetinaNet, and QIR-Net. Higher diagonal values are seen at RetinaNet and QIR-Net compared to other models.
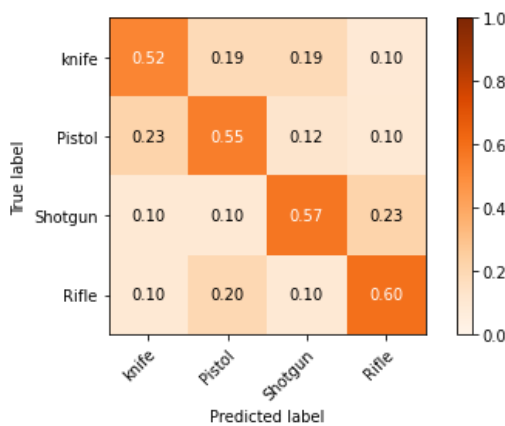

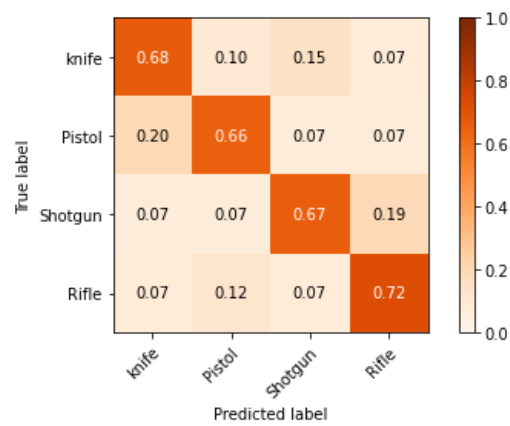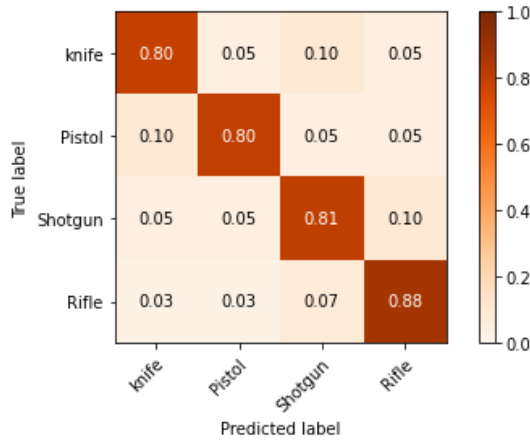Figure 13. Confusion matrix of LeNet   Figure 14. Confusion matrix of AlexNet

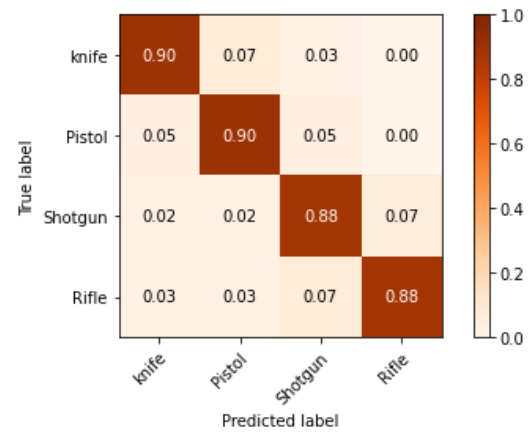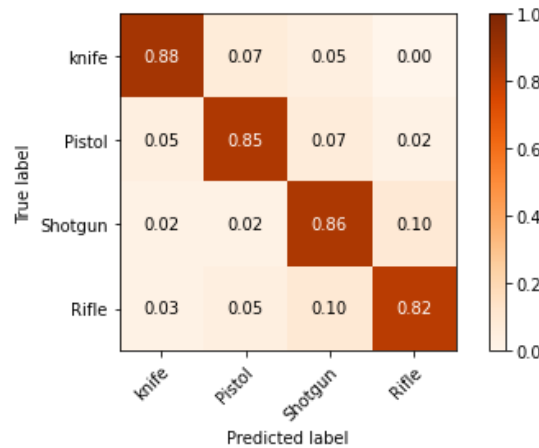Figure 15. Confusion matrix of VGG   Figure 16. Confusion matrix of RetinaNet

Figure 17. Confusion matrix of QIR-Net

4.3. Precision

Precision is another important evaluation feature of the model while measuring performance. It shows the consistency and reliability of the system. In deep learning models, it is the relationship between the number of true positive outcomes and the total positive outcomes, that is, true positive and false positive.

4.4. Recall

The ratio between the true positive outcomes to the sum of true positive and false negative outcomes is termed recall. Recall shows the capability of a model of detecting positive outcomes if it shows more positive outcomes which means there is a higher recall value. A recall is only concerned with positive outcomes and is free from negative outcomes.

4.5. F1-Score

When the model results in numerous incorrect positive orderings or rare accurate positive arrangements, this increases the denominator, and thus reduces the precision. All along with other tools it also helps us to measure the performance of the said model, while comparing with other models. While precision increases when the model results in numerous accurate positive arrangements (maximizing true positive). The model results from less accurate positive classifications (minimize false positive). Equation 4 shows the formula for calculating the F1 score. Figures 18 to 22 show the comparison of the precision, recall, and F1 scores of the aforementioned models. Here also the dominance of QIR-Net and RetinaNet is observed.

Figure 18. Recall, F1 score, and precision of LeNet

Figure 19. Recall, F1 score, and precision of AlexNet

Figure 20. Recall, F1 score, and precision of VGG

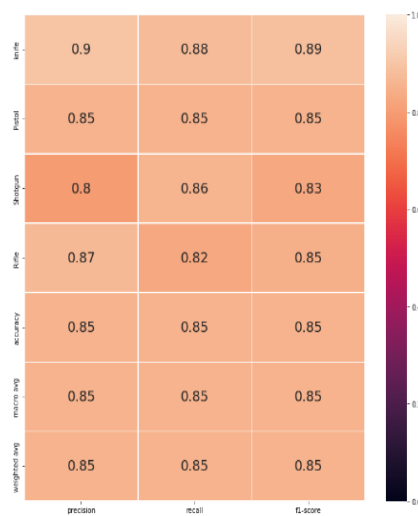Figure 21. Recall, F1 score, and precision of RetinaNet

Figure 22. Recall, F1 score, and precision of QIR-Net

4.6. Reciever operating characteristic

It is a graph that explains the analytical capacity of a binary classifier system as its judgment limit is altered. It is plotted using true +ve versus false +ve in various situations. The receiver operating characteristic (ROC) graphs of LeNet, AlexNet, VGG, QIR-Net, and Retina-Net are shown in Figures 23 to 27.
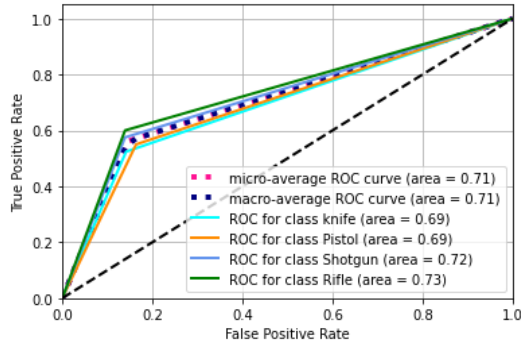
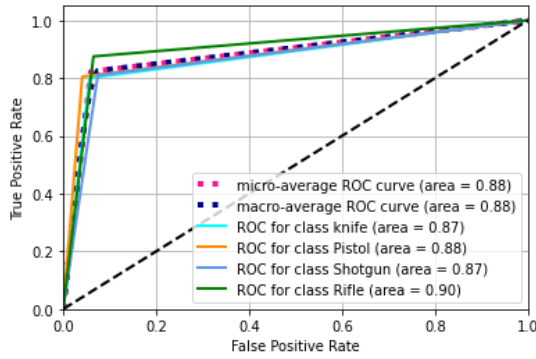Figure 23. ROC of LeNet
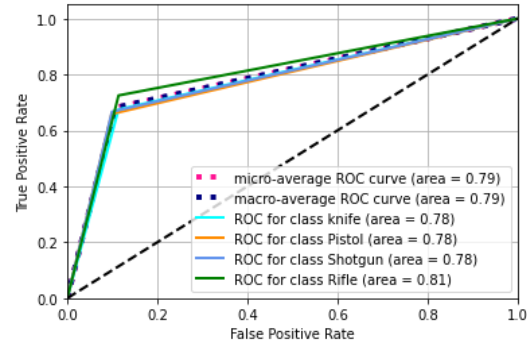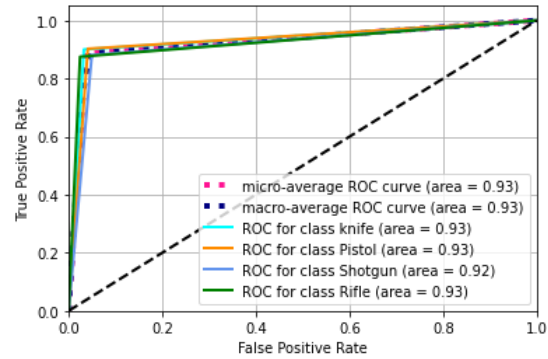
Figure 24. ROC of AlexNet

Figure 25. ROC of VGG
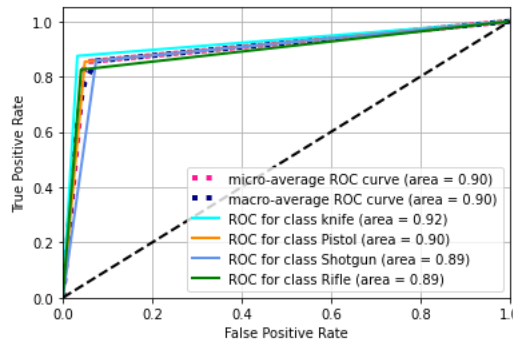
Figure 26. ROC of RetinaNet

Figure 27. ROC of QIR-Net

4.7. Prediction (samples)

Figure 28 presents some sample images for weapon detection using our proposed models. The data collected in different fields are used to verify the operation of the model. These results are taken from movie clips, drama, online footage, and randomly taken images from daily life. The system can detect even smaller arms. From some of the images shown above, one can see the performance of the model. In some cases, the weapon is too small to be seen, but the system has succeeded in detecting it, even though only part of the weapon is visible.
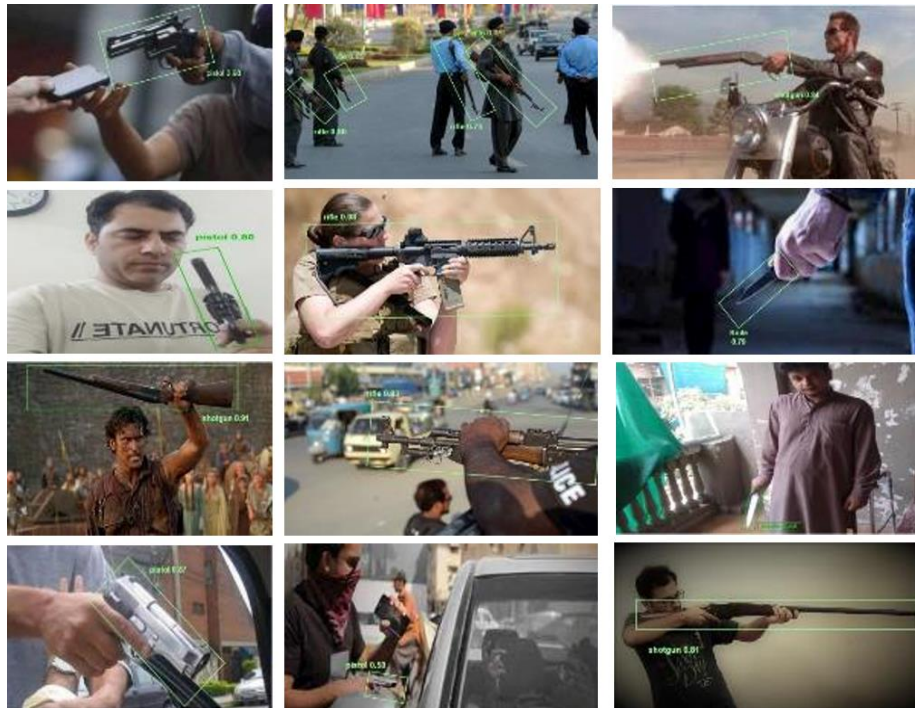
Figure 28. Sample results

5. CONCLUSIONS

As presented in the above research; the rate of street crimes especially in underdeveloped countries is getting higher day by day. Every day violent incidents are registered, which took place mostly because a person has access to the armory. Classical and manual techniques are not efficient enough to meet current constraints. Prevention is the only solution to such dilemmas; otherwise, a cure is not possible. To circumvent such situations a system is required, that is capable of handling the situation before it occurs. A system that informs the concerns as it detects the weapon. So precautionary measures would be taken in advance to abort the incident before it occurs. In this research, the "SMARTED", an intelligent machine is developed to perform automatic surveillance on common public places (especially streets) and inform security when an armory is detected to protect people from any violent behavior. It is observed that previously developed models like FRCNN (family) have shown remarkable accuracy on the images but are unable to perform in real-time scenarios. However, SMARTED is proficient enough to work in real-time situations with a higher accuracy level. To substantiate performance, results are compared with existing models, and a significant improvement is observed. Results through accuracy, F1 score, precision, recursive operative characteristic, and confusion matrix have revealed the efficacy of the model. These tools have deliberately revealed the performance of the RetinaNet and QIR-Net compared to other state-of-the-art models. Another fascinating feature of the paper is to address the solution in conventional computing technology as well as quantum computing technology. For prospects, these models are further tuned to optimize the results; especially QIR-Net can get optimized with fewer changes. Second, these models can be trained on a larger dataset that includes heavy weapons. Therefore, the models could detect weaponry such as tanks, cannons, long-range rifles and hand grenades. These models can also be used in other fields of life to detect objects where precision and speed are the most important concerns.

REFERENCES

[1]   S. A. A. Shah, M. A. Al-Khasawneh, and M. I. Uddin, "Review of weapon detection techniques within the scope of street-crimes," in *2021 2nd International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*, 2021, doi: 10.1109/ICSCEE50312.2021.9498007.
[2]   A. A. Alhussan, A. Al-Dhaqm, W. M. S. Yafooz, A. H. M. Emara, S. B. A. Razak, and D. S. Khafaga, "A Unified Forensic Model Applicable to the Database Forensics Field," *Electron.*, vol. 11, no. 9, pp. 1–20, 2022, doi: 10.3390/electronics11091347.

[3] S. A. A. Shah, M. A. Al-Khasawneh, and M. I. Uddin, "Street-crimes Modelled Arms Recognition Technique (SMART) : Using VGG," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, in *2021 2nd International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*, 2021, doi: 10.1109/ICSCEE50312.2021.9497928.

[4] A. Pesah, M. Cerezo, S. Wang, T. Volkoff, A. T. Sornborger, and P. J. Coles, "Absence of Barren Plateaus in Quantum Convolutional Neural Networks," *Phys. Rev. X*, vol. 11, no. 4, p. 41011, 2021, doi: 10.1103/PhysRevX.11.041011.

[5] S. A. A. Shah, I. Uddin, F. Aziz, S. Ahmad, M. A. Al-Khasawneh, and M. Sharaf, "An Enhanced Deep Neural Network for Predicting Workplace Absenteeism," *Complexity*, vol. 2020, 2020, doi: 10.1155/2020/5843932.

[6] A. Agurto, Y. Li, G. Y. Tian, N. Bowring, and S. Lockwood, "A review of concealed weapon detection and research in perspective," *2007 IEEE Int. Conf. Networking, Sens. Control. ICNSC'07*, no. April, pp. 443–448, 2007, doi: 10.1109/ICNSC.2007.372819.

[7] J. Ruiz-Santaquiteria, A. Velasco-Mata, N. Vallez, G. Bueno, J. A. Alvarez, and O. Deniz, "Handgun detection using combined human pose and weapon appearance," *IEEE Access*, vol. 9, pp. 123815-123826, Sep. 2021, doi: 10.1109/ACCESS.2021.3110335.

[8] H. Muchiri, I. Ateya, and G. Wanyembi, "Concealed Firearm Detection in Male and Female on Video using Machine Learning Classification : A Comparative Study," *International Research Journal of Advanced Engineering and Science*, vol. 4, no. 1, pp. 187–194, 2019.

[9] R. Alhejaili, E. S. Alhazmi, A. Alsaeedi, and W. M. Yafooz, "Sentiment Analysis of The Covid-19 Vaccine For Arabic Tweets Using Machine Learning," in *2021 9th International Conference on Reliability, Infocom Technologies and Optimization*, Sep. 2021, doi: 10.1109/ICRITO51393.2021.9596517,

[10] F. Pérez-Hernández, S. Tabik, A. Lamas, R. Olmos, H. Fujita, and F. Herrera, "Object Detection Binary Classifiers methodology based on deep learning to identify small objects handled similarly: Application in video surveillance," *Knowledge-Based Syst.*, vol. 194, p. 105590, 2020, doi: 10.1016/j.knosys.2020.105590.

[11] A. Warsi, M. Abdullah, M. N. Husen, M. Yahya, S. Khan, and N. Jawaid, "Gun Detection System Using Yolov3," in *2019 IEEE 6th Int. Conf. Smart Instrumentation, Meas. Appl. ICSIMA 2019*, no. May 2020, pp. 1–4, 2019, doi: 10.1109/ICSIMA47653.2019.9057329.

[12] Z. A. Khan *et al.*, "Optimal policy learning for disease prevention using reinforcement learning," *Sci. Program.*, vol. 2020, p. 26000, 2020, doi: 10.1155/2020/7627290.

[13] R. F. Alhujaili and W. M. Yafooz, "Sentiment Analysis for YouTube Educational Videos Using Machine and Deep Learning Approaches," in *2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI)*, May 2022, pp. 238-244, doi: 10.1109/ICETCI55101.2022.9832284.

[14] M. I. Uddin, S. A. A. Shah, M. A. Al-Khasawneh, A. A. Alarood, and E. Alsolami, "Optimal policy learning for COVID-19 prevention using reinforcement learning," *J. Inf. Sci.*, vol. 48, no. 3, pp. 336–348, 2022, doi: 10.1177/0165551520959798.

[15] S. Gawade, R. Vidhya, and R. Radhika, "Automatic Weapon Detection for Surveillance Applications," *SSRN Electron. J.*, 2022, doi: 10.2139/ssrn.4143822.

[16] J. W. Wu, W. Cai, S. M. Yu, Z. L. Xu, and X. Y. He, "Optimized visual recognition algorithm in service robots," *Int. J. Adv. Robot. Syst.*, vol. 17, no. 3, pp. 1–11, 2020, doi: 10.1177/1729881420925308.

[17] R. F. D. A. Kanehisa and A. D. A. Neto, "Firearm Detection using Convolutional Neural Networks," in *ICAART 2019 - Proc. 11th Int. Conf. Agents Artif. Intell.*, 2019, vol. 2, no. January, pp. 707–714, doi: 10.5220/0007397707070714.

[18] M. I. Uddin *et al.*, "Prediction of Future Terrorist Activities Using Deep Neural Networks," *Complexity*, vol. 2020, 2020, doi: 10.1155/2020/1373087.

[19] R. Alhejaili, A. Alsaeedi, and W. Yafooz, "Detecting Hate Speech in Arabic Tweets During COVID-19 Using Machine Learning Approaches," in *Proceedings of Third Doctoral Symposium on Computational Intelligence*, 2023, pp. 467-475, doi: 10.1007/978-981-19-3148-2_39.

[20] T. Hao, X. Li, Y. He, F. L. Wang, and Y. Qu, "Recent progress in leveraging deep learning methods for question answering," *Neural Comput. Appl.*, vol. 34, no. 4, pp. 2765–2783, Feb. 2022, doi: 10.1007/S00521-021-06748-3.

[21] Y. Elmir, S. A. Laouar, and L. Hamdaoui, "Deep learning for automatic detection of handguns in video sequences," *CEUR Workshop Proc.*, vol. 2351, no. May, 2019.

[22] W. Yafooz, A. H. M. Emara, and M. Lahby, "Detecting Fake News on COVID-19 Vaccine from YouTube Videos Using Advanced Machine Learning Approaches," *Combating Fake News with Computational Intelligence Techniques*, pp. 421-435, 2022, doi: 10.1007/978-3-030-90087-8_21

[23] I. Uddin, S. Atif, A. Shah, and M. A. Al-khasawneh, "A novel deep convolutional neural network model to monitor people following guidelines to avoid COVID-19," *Journal of Sensors*, pp. 1–15, 2020, doi: 10.1155/2020/8856801.

[24] S. A. S. A. A. Shah, M. A. Al-Khasawneh, and M. I. I. Uddin, "Street-crimes modelled arms recognition technique (SMART): Using VGG," in *2021 2nd Int. Conf. Smart Comput. Electron. Enterp. Ubiquitous, Adapt. Sustain. Comput. Solut. New Norm. ICSCEE 2021*, 2021, pp. 38–44, doi: 10.1109/ICSCEE50312.2021.9497928.

[25] G. K. Verma and A. Dhillon, "A Handheld Gun Detection using Faster R-CNN Deep Learning," *ACM Int. Conf. Proceeding Ser.*, pp. 84–88, 2017, doi: 10.1145/3154979.3154988.

[26] A. Chalumuri, R. Kune, and B. S. Manoj, "Training an Artificial Neural Network Using Qubits as Artificial Neurons: A Quantum Computing Approach," *Procedia Comput. Sci.*, vol. 171, no. 2019, pp. 568–575, 2020, doi: 10.1016/j.procs.2020.04.061.

[27] A. H. Ashraf *et al.*, "Weapons detection for security and video surveillance using CNN and YOLO-V5s," *Comput. Mater. Contin.*, vol. 70, no. 2, pp. 2761–2775, 2022, doi: 10.32604/cmc.2022.018785.

[28] S. Lu, "Deep learning for object detection in video," *J. Phys. Conf. Ser.*, vol. 1176, no. 4, 2019, doi: 10.1088/1742-6596/1176/4/042080.

[29] L. Pang, H. Liu, Y. Chen, and J. Miao, "Real-time concealed object detection from passive millimeter wave images based on the YOLOV3 algorithm," *Sensors (Switzerland)*, vol. 20, no. 6, 2020, doi: 10.3390/s20061678.

[30] A. Lamas *et al.*, "Human pose estimation for mitigating false negatives in weapon detection in video-surveillance," *Neurocomputing*, vol. 489, pp. 488–503, 2022, doi: 10.1016/j.neucom.2021.12.059.

[31] N. Mast *et al.*, "Channel Contention-Based Routing Protocol for Wireless Ad Hoc Networks," *Complexity*, vol. 2021, 2021, doi: 10.1155/2021/2051796.

[32] A. Castillo, S. Tabik, F. Pérez, R. Olmos, and F. Herrera, "Brightness guided preprocessing for automatic cold steel weapon detection in surveillance videos with deep learning," *Neurocomputing*, vol. 330, no. November, pp. 151–161, 2019, doi: 10.1016/j.neucom.2018.10.076.

[33] C. Y. Wang, H. Y. M. Liao, Y. H. Wu, P. Y. Chen, J. W. Hsieh, and I. H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2020-June, pp. 1571–1580, 2020, doi: 10.1109/CVPRW50498.2020.00203.

[34] T. D. Piyadasa, "Concealed Weapon Detection Using Convolutional Neural Networks," pp. 0–2, 2020.

[35] F. Gelana and A. Yadav, "Firearm Detection from Surveillance Cameras Using Image Processing and Machine Learning Techniques," in *Smart Innovations in Communication and Computational Sciences: Proceedings of ICSICCS-2018*, 2019, pp. 25-34, doi: 10.1007/978-981-13-2414-7_3.

[36] V. V. Danilov *et al.*, "Real-time coronary artery stenosis detection based on modern neural networks," *Sci. Rep.*, vol. 11, no. 1, pp. 1–14, 2021, doi: 10.1038/s41598-021-87174-2.

[37] J. L. S. González, C. Zaccaro, J. A. Álvarez-García, L. M. S. Morillo, and F. S. Caparrini, "Real-time gun detection in CCTV: An open problem," *Neural Networks*, vol. 132, no. October, pp. 297–308, 2020, doi: 10.1016/j.neunet.2020.09.013.

[38] S. Shi *et al.*, "Quantum-inspired complex convolutional neural networks," *Appl. Intell.*, 2022, doi: 10.1007/s10489-022-03525-0.

[39] D. Mattern, D. Martyniuk, H. Willems, F. Bergmann, and A. Paschke, "Variational Quanvolutional Neural Networks with enhanced image encoding," no. June, 2021, [Online]. Available: http://arxiv.org/abs/2106.07327.

[40] S. M. R. Taha and A. K. Nawar, "A New Quantum Radial Wavelet Neural Network Model Applied to Analysis and Classification of EEG Signals," *Int. J. Comput. Appl.*, vol. 85, no. 7, pp. 7–11, 2014, doi: 10.5120/14851-3216.

[41] I. Cong, S. Choi, and M. D. Lukin, "Quantum convolutional neural networks," *Nat. Phys.*, vol. 15, no. 12, pp. 1273–1278, 2019, doi: 10.1038/s41567-019-0648-8.

[42] T. Hur, L. Kim, and D. K. Park, "Quantum convolutional neural network for classical data classification," *Quantum Mach. Intell.*, vol. 4, no. 1, 2022, doi: 10.1007/s42484-021-00061-x.

[43] O. Lockwood and M. Si, "Reinforcement Learning with Quantum Variational Circuits Reinforcement Learning with Quantum Variational Circuits," in *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 2020, vol. 16, no. 1, pp. 245-251, doi: 10.1609/aiide.v16i1.7437.

[44] A. N. Hassan, M. J. Mohsin, and A. H. K. Khwayyir, "Discrimination of ECG signal based on S-interpolation and quantum neural network," *Ann. Trop. Med. Public Heal.*, vol. 23, no. 10, 2020, doi: 10.36295/ASRO.2020.231039.

[45] S. Wei, Y. Chen, Z. Zhou, and G. Long, "A quantum convolutional neural network on NISQ devices," *AAPPS Bull.*, vol. 32, no. 1, pp. 22–24, 2022, doi: 10.1007/s43673-021-00030-3.

[46] R. Olmos, S. Tabik, and F. Herrera, "Automatic handgun detection alarm in videos using deep learning," *Neurocomputing*, vol. 275, no. February, pp. 66–72, 2018, doi: 10.1016/j.neucom.2017.05.012.

BIOGRAPHIES OF AUTHORS

Syed Atif Ali Shah 🆔 📷 SC ◐ is a lecturer at the Department of Computer Sciences at Air University, Islamabad, since 2022. From 2016 to 2021 he was serving as an assistant professor at the Northern University. From 2015 to 2016 he served as a Master Trainer (Vendors) and Team Lead (User Acceptance Testing) in Oracle Vision and Automation Project at Mobilink (PMCL). From 2010 to 2014 he served as a senior lecturer and as a HOD at Brains Institute. During 2008-2010 and 2014-2015, he joined the industry as a software engineer. He has also been a visiting faculty member at NUST, FUUAST, ICMS, and the Afghan Institute of Learning. He started his professional carrier in 2004 as a lecturer at CSDC and then in 2005 moved to Abasyn University where he served as a Head of the CS department and a Program Coordinator. Recently pursuing a Ph.D. in Deep Learning from Malaysia. He received his MS in Electrical Engineering from Abasyn University and his MS in Computer Science (specialization in Software Engineering) from Virtual University. In 2004 earned an MSc in Computer Science from the University of Peshawar. His research interests lie in the area of Data Science, Deep Learning, Neural Networks, and Software Engineering. Much of his work is on innovations and renovations for improving the design, mainly by applying neural networks and redesigning models. He has explored the cooperation of computational models in various domains. He has also investigated the insinuation of Deep Learning in predicting actions. He has published numerous articles in international journals and conferences. He also contributed to IEEE Access as a volunteer reviewer. He can be contacted at email: atif.ali@mail.au.edu.pk.

Dr. Abdel-Hamid Emara 🆔 📷 SC ◐ is an associate professor in the Department of Computers and Systems Engineering, Faculty of Engineering, Al-Azhar University, Cairo, Egypt He received his BSc, MSc, and PhD in Computers and Systems Engineering from Al-Azhar University in 1992, 2000, 2006, respectively. He supervised number of students at the master and PhD levels. Furthermore, He delivered and conducted many workshops in the research area He served as a member of various committees in many international conferences. He is a volunteer reviewer with different peer-review journals. He is currently an Assistant Professor in Computer Science Department at University of Taibah at Al-Madinah Al Monawarah, KSA. His research interest includes, Data Mining, Machine Learning, Deep Learning, Natural Language Processing, Social Network Analytics and Data Management. He published several research papers and participated in several in International journal and local/international conferences. She can be contacted at email: aemara@taibahu.edu.sa.

Dr. Nasir Algeelani 🆔 ⑧ SC ⟳ is currently a Senior Lecturer in the Computer Science Department at Al-Madinah International University (MEDIU), Kuala Lumpur, Malaysia. Previously he worked as Lecturer at Industrial Technical Institute (ITI) from 2000 to 2010. Dr. Nasir Ahmed Algeelani received the B.E. degree in electrical power system from University of Aden, Yemen, Aden, in 1997, the M.E. degree in electrical power system engineering from University Technology Malaysia in 2009 and the Ph.D. degree in high voltage engineering from University Technology Malaysia in 2014. He conducted a post-doctorate at high voltage engineering department at University Technology Malaysia (2014-2016). He has published as authored and co-authored more than 30 papers in various technical journals and conference proceedings. His research interests include high-voltage instrumentation, partial discharge, detection and warning systems and condition monitoring of high-power equipment. He can be contacted at email: nasir.ahmed@mediu.edu.my.
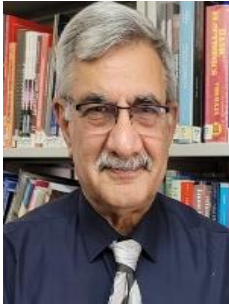
Dr. Najeeb 🆔 ⑧ SC ⟳ jointed MEDIU in SEPT.2012 as a lecturer in Faculty of computer and Information Technology. He completed my M.Sc. from North Staffordshire University in UK, worked in computer center in Iraq for more than 15 years as a Software manager. After completed Ph.D. start working in private University in Iraq. Having over 15 years' experience as senior lecturer, also as a Dean of Private University College, Head of Computer Department for more than 5 years. Since SEPT.2012 in MEDIU had many responsibility and academic positions such as Dean of Library Deanship, Dean of ICT Faculty. Taught many Bachelor and Mater subjects, and still teaching each semester in the area of E-Services. Also working as an Asso Prof. to supervise many Master and Ph.D. students. Issues many research papers and join many international conferences. He can be contacted at email: dr.najeeb@mediu.edu.my.