

Stand-off concealed firearm detection using motion tracking and convolutional neural networks

Henry Muchiri Muriithi¹, Ismail Lukandu Ateya¹, Gregory Wanyembi²

¹School of Computing and Engineering Sciences, Strathmore University, Nairobi, Kenya

²School of Computing and Informatics, Mount Kenya University, Nairobi, Kenya

Article Info

Article history:

Received Mar 23, 2023

Revised Feb 10, 2024

Accepted Feb 24, 2024

Keywords:

Automated concealed firearm detection

Convolutional neural networks

Skeletal-based motion tracking

Smart surveillance

ABSTRACT

The standoff detection of concealed firearms is crucial in managing public security in public spaces. Currently employed standoff concealed weapon detection techniques employ electromagnetic wave imaging which has been found to be extremely slow and may require expensive hardware and may not be applicable in public open spaces. In order to maintain safety in open spaces, artificial intelligence enabled video surveillance systems have been widely adopted. This poses an opportunity to explore video surveillance cameras as concealed weapon detectors. A review of existing video surveillance based automated weapon detection approaches discovered that the focus was on the detection of unconcealed firearms leaving a gap in the detection of concealed firearms. This study addresses the aforementioned gap by providing a standoff concealed firearm detection approach on video based on skeletal-based human motion tracking and convolutional neural networks (CNN). The motion of armed and unarmed persons was tracked using a depth camera and further classified using CNN model. The developed model reported 100% accuracy, precision, and recall scores. These results outperformed results obtained from traditional machine learning models therefore highlighting the superior capability of the proposed approach for concealed firearm detection on video to complement the efforts of human video surveillance operators.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Henry Muchiri Muriithi

School of Computing and Engineering Sciences, Strathmore University

Ole Sangale Rd, off Langata Road, Madaraka Estate, Nairobi, Kenya

Email: hmuchiri@strathmore.edu

1. INTRODUCTION

All over the world, populations are facing an increasing burden of firearm violence and mass shootings on mortality and injuries [1]. Mortality from firearms contributes more than 500 deaths each day worldwide with a majority of these deaths resulting from homicides [1]. An additional 2,000 people are injured or maimed by gunshots every day [2]. With these alarming statistics, providing more significant control over firearm usage is a crucial factor in reducing the effects of firearm violence. A challenging task for law enforcement officers is the detection of concealed firearms [3]. Concealed weapon detection approaches can broadly be categorized as either stop and search approaches or standoff approaches [3]. Stop and search approach for example walk through metal detectors requires persons to stop at screening stations to be searched while standoff approaches suspects are screened at a distance [3]. Stop and search approaches are however only limited to entry points of buildings and therefore leaving out other areas such as open streets where the public is also in danger of attacks from firearms. Standoff approaches on the converse can be employed in open streets.

Among the solutions employed for the standoff concealed weapon detection is the application of electromagnetic wave imaging techniques or the use of video surveillance [3], [4]. The electromagnetic techniques include the use of ultrasound, mmWave, Terahertz, infrared, fusion of visual RGB image and infrared and X-ray imaging [5]. Various researchers have proposed electromagnetic based standoff concealed weapon detection solutions [5]–[8]. Some researchers concur that the application of electromagnetic wave imaging techniques have a long processing time and require expensive hardware that may not be possible in many places such as open streets [4], [5].

Currently there are millions of video surveillance cameras installed in many open streets to maintain security [9]. These systems rely on human operators to manna them and communicate to officers on the ground incase suspicious activity or behavior is observed [9], [10]. Video surveillance systems however, have the limitation that they require constant human supervision which is impractical especially in vast volumes of data [4], [11]. An attractive alternative is the deployment of automated video surveillance systems where potential criminal activites can be autonomously detected using artificial intelligence techniques and prevent them before they occure [3], [10]. Various authors have proposed solutions for intelligent standoff weapon detection techniques on video surveillance systems.

Ahmed *et al.* [9] implemented a real-time weapon detection approach using a scaled YOLOv4 object detector with the ability to detect unconcealed firearms with high mean average precision rates of over 92%. The approach achieved lower latency, higher throughput, and improved privacy by deploying on Jetson Nano edge computing device. Narejo *et al.* [10] proposed a firearm detection technique using YOLOv3 object detection neural network with the ability to detect unconcealed firearms and subsequently sending an alarm to security enforcement personnel. This approach outperformed intitial approaches that employed YOLOv2 and traditional convolutional neural network (CNN) approaches. Bhatti *et al.* [11] developed a firearm detection technique in real-time CCTV using YOLOV4 object detection neural network that could detect unconcealed firearms in low resolution and brightness with over 91% average precision and F1 scores. Figure 1 illustrates a sample detection outcome from the approach.



Figure 1. Demonstration of the detection of a unconcealed firearm by [11]

Sumi and Dey [12] developed an unconcealed firearm detector on video using YOLOv5 object detection neural network. A comparative study revealed the models superior detection ability compared to the baseline faster-RCNN based firearm detector approaches. Additionally, the study revealed that the application of augmented datasets yielded superior performance to non-augmented data.

The review of existing automated standoff firearm detection approaches on video surveillance reveals great success and strides made in the detection of visible/unconcealed firearms. This however presents a fundamental gap in the detection of concealed firearms on video surveillance. The detection of concealed weapons people's clothing is crucial in maintaining public safety [13].

This study aims to address the identified gap by presenting a real time automated standoff concealed handgun detection by applying skeletal-based motion tracking technique to automatically track changes in human motion on video as they walk with a concealed firearm tucked on their hip using state of the art deep learning models. This approach is premised on the findings by [14], [15] which revealed that trained CCTV human operators have the ability to identify concealed firearms because when a firearm is concealed into the trouser pocket or the front waistband, it may hinder leg movements on that side of the body resulting in the right stride being shorter than the left and a shorter arm swing. They attribute this disruption to the individual attempts to either conceal the weapon or limit its movement so as not to drop it.

The proposed standoff surveillance solution would address the existing gap by enabling the detection of concealed firearm on video surveillance while people are in motion for example on streets and allow early detection before they are used to commit a crime. This approach would allow law enforcement

officers to rapidly and reliably screen individuals for concealed threats without any physical contact or significant disruption of the suspect(s) activities and in effect reduce street crimes [5]. Automated motion tracking has successfully been applied in various areas. In [16], [17] tracked motion for personality assessment, [18], [19] tracked motion for gender classification and age estimation, [20], [21] tracked motion for person identification and biometrics among others. To the best of our knowledge, there is no previous work done to deploy motion tracking human pose estimation technique for concealed weapon detection on video. The main contributions of this work are:

- Development of a novel 3D skeletal-based motion tracking dataset containing armed and unarmed participants.
- Presentation of a concealed firearm detection approach on video using human motion tracking technique and CNN techniques.
- Extend our previous study [22] that applied traditional machine learning algorithms for concealed firearm detection.

The rest of the paper is organized as follows. Section 2 elaborates on the methods used to develop the proposed solution. Section 3 contains the presentation of the research results and a discussion that provides the reader with a deeper insight into the research findings. Finally, section 4 concludes the paper and provides possible future directions in this area.

2. METHODS

In this section, detailed description and justification of the applied materials and methods set out to achieve the study objectives.

2.1. Data collection instruments and procedure

The Microsoft Kinect depth camera together with the accompanying Kinect for windows software development kit (SDK) was employed to track participants motion. Specifically, the Kinect v2 was used, which enables capturing images at up to 30 frames per second (fps) and track 25 joints in the human body. The camera captures the 3D spatial information of the tracked object relative to the camera. Microsoft Kinect and other 3D sensing technologies offer real-time skeletal detection and tracking algorithms. Skeleton-based motion tracking was preferred because they have lower data dimensionality in comparison with RGB-based representations and therefore allowing the development of computationally faster and less complex systems [23]. A Ceska handgun loaded with 11 rounds of ammunition was used. The pistol had a weight of 2.77 kilograms (kgs) and a length of 0.2 metres. A handgun was considered in this study due to its rampant use in criminal activities [24].

The dataset creation process involved skeletal tracking and recording the 3D spatial-temporal skeletal motion patterns of participants when armed and when unarmed. This was done in a lab environment on a 6 metres long by 1-meter-wide clearly marked walkway using the Microsoft Kinect depth camera. The sensor was elevated 1.2 metres above the ground and 1.5 metres away from one end of the walking path to ensure that the subjects were within the range and field of view of the sensor for skeletal tracking. A physical marker was placed near the end of the footpath, so that the subjects were aware of where they needed to stop without having to look down to avoid skewing skew the results. Figure 2 illustrates the skeletal tracking of one of the participants walking across the data collection pathway.

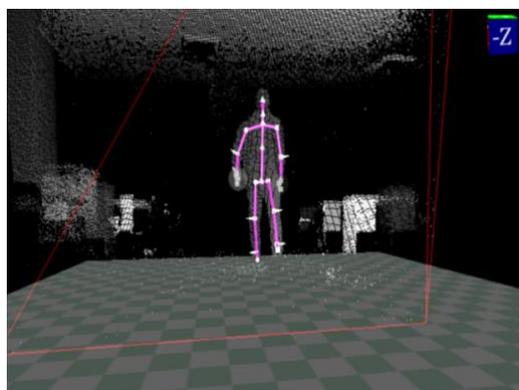


Figure 2. Data collection illustration

To measure the classification performance of the developed model, the study employed various complementary metrics that were captured during model training and during testing and validation phases [27]. The models learning accuracy, and loss function across the training and testing cycle was plotted in order to diagnose the models behaviour. During testing and validation phase the confusion matrix and precision, recall and F1 scores were measured.

Figure 4 provides a comparative plot of the training and validation loss function of the model across the 15 epochs. This plot is beneficial in indicating the overall behavior of the developed neural network. The plot depicts a good fit characterized by a training and validation loss which both decrease to a point of stability with a small gap between the two metrics [27]. This result is an indication that the model is not overfitting or underfitting the data and therefore implying that the developed model was adequately learning and was able to generalize well.

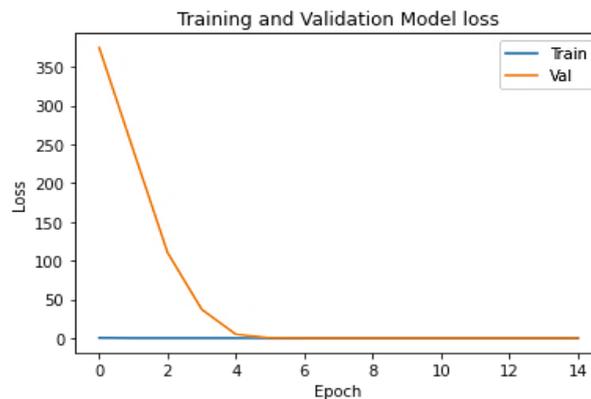


Figure 4. Training and validation loss plot

The developed model reported a training and validation accuracies of 100% over 15 epochs as illustrated by Figure 5. A significantly small gap between the training accuracy and validation accuracy signifies a good model fit which is devoid of under or overfitting. A validation accuracy of 100% which is obtained without overfitting denotes the superior capability of the developed model to accurately classify motion images of armed and unarmed.

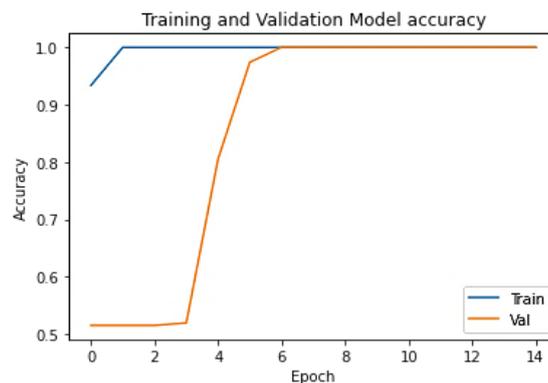


Figure 5. Training and validation accuracy plot

To further evaluate and understand the performance of the developed concealed firearm detection CNN, the study employed the confusion matrix performance measurement tool. The obtained results are presented in Figure 6. The presented results further affirm the superior performance of the model with zero classification errors reported as false positives (FP) or false negatives (FN). This is an indication that the developed model can accurately classify armed data (true positives (TP)) and distinguish that from images of unarmed instances (true negatives (TN)).

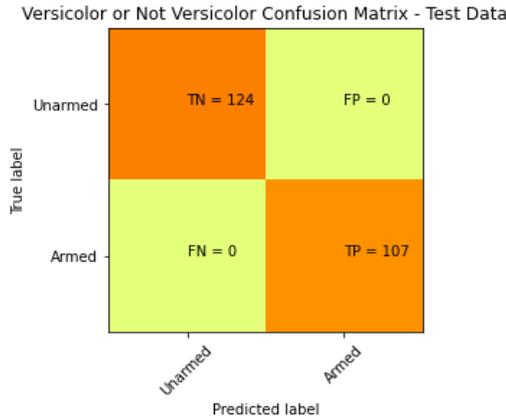


Figure 6. Armed or unarmed confusion matrix

Because the developed model was a classification model, precision and recall performance metrics were additionally employed [27]. The scores were compared with those of our previous study [22] that employed traditional machine learning algorithms. The deep learning model presented 100% scores in all the matrix in comparison to traditional machine learning algorithms which reported a maximum of 93% in all measured matrix. The comparative results are presented in Table 1.

Table 1. Comparative analysis of the results

Algorithms	Evaluation metrics		
	Accuracy (%)	Precision (%)	Recall (%)
Developed CNN	100	100	100
k-nearest neighbour	93.6	93.4	93.6
Random forest	90.6	90.7	90.6
Decision tree	68.60	68.60	68.60
Support vector machine	41.1	40.9	41.1
naïve Bayes	37.5	39.2	37.5

These results further confirm the superiority of the deep learning approach. The presented precision scores indicate the correctness of the model to classify an image as armed with no cases of FP while the presented recall scores indicate that the neural network’s ability to correctly classify all positive/armed cases which in this case are the armed images. Additionally, the study was keen on measuring the detection time. The video data used to train this model contained 81 frames recorded in about 4 seconds for a distance of about 6 meters. The time of 4 seconds can be interpreted as the average walking time required for the concealed firearm detection model to analyze motion and detect the firearm. This time is short and acceptable for the detection task.

4. CONCLUSION

This study aimed to develop an efficient deep learning model for the standoff detection of people carrying concealed firearms by tracking their motion. The model’s learning curve was evaluated together with classification performance metrics such as accuracy, precision, and recall scores. The analysis of the learning curve shows a good fit learning curve which indicates the model is able to satisfactorily learn the distinguishing features in the image inputs. The model presented an accuracy, precision, and recall scores of 100%. This finding is a confirmation of the ability of the developed CNN model to accurately classify armed and unarmed images by 3D motion analysis. Therefore, this research concludes that it is possible to a great extent to apply automated motion analysis using state-of-the-art video analysis techniques for concealed firearm detection. This approach to automated concealed firearm detection by motion analysis is novel and a great breakthrough in this area and will go along way in aiding the fight against crimes involving firearms. The data applied in the study consisted of single persons walking towards the detection sensor/camera. Following the outstanding performance presented by the developed model, future studies can apply data consisting of multiple persons walking in different directions from the detection sensor/camera. This extension to the study will be a great effort in mimicking real-life surveillance environments. Additionally,

the study was focused in an indoor lab environment, and as such future studies can focus on application of the approach in real-time cctv footage.

ACKNOWLEDGEMENTS

We thank the National Research Fund (NRF) Kenya for the research grant- [Postgraduate grant 2016/2017], Kenya Police Service for providing the firearm that was used in this study, the 27 study participants for their assistance in data acquiring and research assistants for their support in cleaning and pre-processing the data.

REFERENCES

- [1] M. Werbick, I. Bari, N. Paichadze, and A. A. Hyder, "Firearm violence: a neglected 'global health' issue," *Globalization and Health*, vol. 17, no. 1, 2021, doi: 10.1186/s12992-021-00771-8.
- [2] I. Cardoza, J. P. G. -Vázquez, A. D. -Ramírez, and V. Q. -Rosas, "Convolutional neural networks hyperparameter tuning for classifying firearms on images," *Applied Artificial Intelligence*, vol. 36, no. 1, Apr. 2022, doi: 10.1080/08839514.2022.2058165.
- [3] A. Agurto, Y. Li, G. Y. Tian, N. Bowring, and S. Lockwood, "A review of concealed weapon detection and research in perspective," in *2007 IEEE International Conference on Networking, Sensing and Control, ICNSC'07*, IEEE, 2007, pp. 443–448, doi: 10.1109/ICNSC.2007.372819.
- [4] 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: Ubiquitous, Adaptive, and Sustainable Computing Solutions for New Normal, ICSCCE 2021*, 2021, pp. 26–37, doi: 10.1109/ICSCCE50312.2021.9498007.
- [5] X. Gao, H. Liu, S. Roy, G. Xing, A. Alansari, and Y. Luo, "Learning to detect open carry and concealed object with 77 GHz radar," *IEEE Journal on Selected Topics in Signal Processing*, vol. 16, no. 4, pp. 791–803, 2022, doi: 10.1109/JSTSP.2022.3171168.
- [6] Z. Zhang, X. Di, Y. Xu, and J. Tian, "Concealed dangerous object detection based on a 77GHz radar," in *2018 IEEE International Workshop on Electromagnetics: Applications and Student Innovation Competition, iWEM 2018*, IEEE, 2018, doi: 10.1109/iWEM.2018.8536660.
- [7] Y. Li, Z. Peng, R. Pal, and C. Li, "Potential active shooter detection based on radar micro-doppler and range-doppler analysis using artificial neural network," *IEEE Sensors Journal*, vol. 19, no. 3, pp. 1052–1063, 2019, doi: 10.1109/JSEN.2018.2879223.
- [8] C. V. Nelson, "Wide-area metal detection system for crowd screening," in *Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Defense and Law Enforcement*, vol. 5071, Sep. 2003, pp. 380–387, doi: 10.1117/12.484846.
- [9] S. Ahmed, M. T. Bhatti, M. G. Khan, B. Lövrström, and M. Shahid, "Development and optimization of deep learning models for weapon detection in surveillance videos," *Applied Sciences*, vol. 12, no. 12, 2022, doi: 10.3390/app12125772.
- [10] S. Narejo, B. Pandey, D. E. Vargas, C. Rodriguez, and M. R. Anjum, "Weapon detection using YOLO V3 for smart surveillance system," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–9, 2021, doi: 10.1155/2021/9975700.
- [11] M. T. Bhatti, M. G. Khan, M. Aslam, and M. J. Fiaz, "Weapon detection in real-time CCTV videos using deep learning," *IEEE Access*, vol. 9, pp. 34366–34382, 2021, doi: 10.1109/ACCESS.2021.3059170.
- [12] L. Sumi and S. Dey, "YOLOv5-based weapon detection systems with data augmentation," *International Journal of Computers and Applications*, vol. 45, no. 4, pp. 288–296, 2023, doi: 10.1080/1206212X.2023.2182966.
- [13] 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*, vol. 20, no. 6, 2020, doi: 10.3390/s20061678.
- [14] I. T. Darker, A. G. Gale, and A. Blechko, "CCTV as an automated sensor for firearms detection: human-derived performance as a precursor to automatic recognition," in *Unmanned/Unattended Sensors and Sensor Networks V*, 2008, doi: 10.1117/12.800264.
- [15] N. Meehan, C. Strange, and A. Garinther, "It's the walk, not the talk: behavioral indicators of concealed and unholstered firearms carrying," *Police Journal*, vol. 94, no. 4, pp. 462–480, Sep. 2021, doi: 10.1177/0032258X20960777.
- [16] J. Sun *et al.*, "Relationship between personality and gait: Predicting personality with gait features," in *Proceedings - 2018 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018*, IEEE, 2019, pp. 1227–1231, doi: 10.1109/BIBM.2018.8621300.
- [17] L. Satchell, P. Morris, C. Mills, L. O'Reilly, P. Marshman, and L. Akehurst, "Evidence of big five and aggressive personalities in gait biomechanics," *Journal of Nonverbal Behavior*, vol. 41, no. 1, pp. 35–44, Sep. 2017, doi: 10.1007/s10919-016-0240-1.
- [18] C. Xu *et al.*, "Real-time gait-based age estimation and gender classification from a single image," in *Proceedings - 2021 IEEE Winter Conference on Applications of Computer Vision, WACV 2021*, IEEE, 2021, pp. 3459–3469, doi: 10.1109/WACV48630.2021.00350.
- [19] L. Igual, À. Lapedriza, and R. Borràs, "Robust gait-based gender classification using depth cameras," *Eurasip Journal on Image and Video Processing*, vol. 2013, no. 1, 2013, doi: 10.1186/1687-5281-2013-1.
- [20] H. Yamada, J. Ahn, O. M. Mozos, Y. Iwashita, and R. Kurazume, "Gait-based person identification using 3D LiDAR and long short-term memory deep networks," *Advanced Robotics*, vol. 34, no. 18, pp. 1201–1211, 2020, doi: 10.1080/01691864.2020.1793812.
- [21] C. Á. -Aparicio, Á. M. G. -Higueras, M. Á. G. -Santamarta, A. C. -Vega, V. Matellán, and C. F. -Llamas, "Biometric recognition through gait analysis," *Scientific Reports*, vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-18806-4.
- [22] H. M. Muriithi, I. A. Lukandu, and G. W. Wanyembi, "Determining the location of a concealed handgun on the human body using marker-less gait analysis and machine learning," in *2nd International Conference on Next Generation Computing Applications 2019, NextComp 2019*, IEEE, pp. 1–6, Sep. 2019, doi: 10.1109/NEXTCOMP.2019.8883635.
- [23] V. Silva, F. Soares, C. P. Leão, J. S. Esteves, and G. Vercelli, "Skeleton driven action recognition using an image-based spatial-temporal representation and convolution neural network," *Sensors*, vol. 21, no. 13, 2021, doi: 10.3390/s21134342.
- [24] K. Nettles, C. Ford, and P. A. Prada-Tiedemann, "Development of profiling methods for contraband firearm volatile odor signatures," *Frontiers in Analytical Science*, vol. 1, 2022, doi: 10.3389/frans.2021.785271.
- [25] Y. Zhu, Y. Zhao, and S. C. Zhu, "Understanding tools: task-oriented object modeling, learning and recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, 2015, pp. 2855–2864, doi:

10.1109/CVPR.2015.7298903.

- [26] H. H. Pham, L. Khoudour, A. Crouzil, P. Zegers, and S. A. Velastin, "Exploiting deep residual networks for human action recognition from skeletal data," *Computer Vision and Image Understanding*, vol. 170, pp. 51–66, 2018, doi: 10.1016/j.cviu.2018.03.003.
- [27] M. Z. Naser and A. H. Alavi, "Error metrics and performance fitness indicators for artificial intelligence and machine learning in engineering and sciences," *Architecture, Structures and Construction*, vol. 3, no. 4, pp. 499–517, Nov. 2023, doi: 10.1007/s44150-021-00015-8.

BIOGRAPHIES OF AUTHORS



Dr. Henry Muchiri Muriithi    holds a Ph.D. in Information Systems from Strathmore university. He is Currently a lecturer and the Academic Director at the Strathmore University School of Computing and Engineering Sciences (SCES). He has been involved in several research projects notably the Microsoft Wise4Africa Project 2019 and the Clinton Health Access Initiative (CHAI) – Drug Access Project 2015. He is also a reviewer for conferences and journals. He has presented conference papers and published articles in various national and international reputable journals of information systems and artificial intelligence. He can be contacted at email: hmuchiri@strathmore.edu.



Prof. Ismail Lukandu Ateya    Professor Ismail Ateya earned his Doctor of Science degree in Applied Geophysics from Kyoto University, Japan in 2003 and Graduate Diploma in Computer Science from University of Auckland, New Zealand in 2006. He is currently the Director, Office of Faculty Affairs at Strathmore University and lecturer at Strathmore University School of Computing and Engineering Science. His research interests are on software and database modelling of large information systems – and he has been an active member of the Database Research Group. He can be contacted at email: iateya@strathmore.edu.



Prof. Gregory Wanyembi    holds a doctorate in management of ICTs from Delft University of Technology, The Netherlands, in 2002. He has published widely in the area of management and strategic utilization of ICTs in developing countries. He is engaged in several research projects and supervision of postgraduate students at both masters and Ph.D. from different universities in Kenya. He is a reviewer of journal articles. He is professor at Mount Kenya University, Kenya. He can be contacted at email: g.wanyembi@gmail.com.