Bushfire Risk Detection Using Internet of Things: An Application Scenario

Mohammad Reza Nosouhi, *Member, IEEE*, Keshav Sood, Neeraj Kumar, *Senior Member, IEEE*, Tricia Wevill, and Chandra Thapa, *Member, IEEE*

***Abstract*—With rising temperatures and events contributing to climate change, the world is facing extreme weather patterns. Recently, Australia was hit hard by bushfires, the most devastating fires ever faced by the country. The economic damage reported was nearly one billion Australian dollars and an estimated three billion native animals were killed or adversely affected. Given the extent and intensity of this damage, researchers are seeking effective solutions to enable the prediction of fire before it starts to increase the time available for firefighters to protect lives and assets and prepare to mitigate the fires. This motivated us to investigate an approach to address this critical problem. In this article, we propose a machine learning (ML)-based approach that detects anomalies in spatiotemporal measurements of environmental parameters (e.g., temperature, relative humidity, etc.). In the proposed approach, an ML-based model learns the normal spatiotemporal behavior of the environmental data (collected over a period of one year). This is carried out during a one-time training phase. Then, during the detection phase, any spatiotemporal pattern in the real-time data (received from the field sensors) that is different than the normal pattern will be identified by the model as anomaly which indicates a possible bushfire situation. Following this, we propose a supplementary classification model based on Moran's I index to ensure that the detected anomalies are not due to either a sensor failure or a security attack (which are common in Internet of Things). We developed three different ML models for performance evaluation and comparison and used the Forest Fire data set to train them. The results of our experiments confirm the effectiveness of the proposed approach in the early detection of fire symptoms.**

***Index Terms*—Agriculture 4.0, artificial intelligence, Australian bushfire, edge computing, Internet of Things (IoT).**

I. Introduction and Background

BUSHFIRE is commonly known as an unplanned vegetation fire or wildfire (grass fires, forest fires, and scrub fires, etc.), at a large scale. This occurs due to both natural and human–intervention related factors, such as weather, natural phenomena, topography, carbon emissions, seasonality, etc. [1]. On one side, essentially, it is shaping the nature and planet naturally in a variety of ways, but, on the other hand, the high frequency of occurrence of bushfire adversely impacts flora and fauna, both human and animal life, damage to properties, creates health issues due to air pollution, post bushfire psychological problems, and high economic impacts. Many areas in Australia are prone to bushfire as the climate is generally dry and hot in many states [2] and drought is common. For natural hazards and disasters, including bushfires, Geoscience Australia is responsible for preparedness and management activities. It has been reported that the bushfire season 2019–2020 was the worst season recorded in Australia [3], [4]. Some devastating impacts reported in [3] and [5], caused by the 2019–2020 bushfire season are listed as follows.

1) More than one billion animals died and three billion were affected in total, which is the highest rate of species loss in any area in the world [6].
2) The air pollution caused by severe bushfires in 2020. The Australian Capital Territory (ACT), Canberra measured the worst air quality index of any major city in the world in January the same year. This has adversely affected territory's residents with severe health risks, including [7] asthma, impaired vision, increased risk of heart attacks, and the development of neurological conditions, etc.
3) This has significantly affected the farming and tourism industry. Severe economic damage was reported in 2019–2020 due to the Australian bushfire, nearly one billion Australian dollars had been reported as insurance losses as a result of the bushfires, almost 13 million hectares of land was been burned. [8].
4) Furthermore the total costs associated with extreme weather and climate change is difficult to quantify. The damage-related loss in property values from climate hazards across Australia in 2030 is estimated at 571 billion Australian dollars [3]. Overall, we note that the impact of bushfire is devastating in Australia.

Towards this problem, it is critical to have effective and accurate forecasting systems to be in place to make better decisions for the mitigation of fire risks and hazards. Many attempts have been made as briefly mentioned below.

Satellites-Based Solutions: In [9]–[11] satellite images are analyzed to detect areas that are generating a high level of infrared radiation (called hot spots). However, this approach suffers from high latency between successive scans, i.e., rescanning of the same area needs a significant amount of

Manuscript received July 10, 2021; accepted August 22, 2021. Date of publication September 6, 2021; date of current version March 24, 2022. *(Corresponding author: Keshav Sood.)*

Mohammad Reza Nosouhi and Keshav Sood are with the Centre of Cyber Security Research Innovation, School of Information Technology, Deakin University, Geelong, VIC 3220, Australia (e-mail: m.nosouhi@deakin.edu.au; keshav.sood@deakin.edu.au).

Neeraj Kumar is with the Department of Computer Science and Engineering, Thapar University, Patiala 147004, India, also with the Department of Computer Science and Information Engineering, Asia University, Taichung City 413, Taiwan, and also with the School of Computer Science, University of Petroleum and Energy Studies, Dehradun 248007, India (e-mail: neeraj.kumar@thapar.edu).

Tricia Wevill is with the School of Life and Environmental Sciences, Deakin University, Geelong, VIC 3220, Australia (e-mail: tricia.wevill@deakin.edu.au).

Chandra Thapa is with the Distributed Systems Security, Data61, CSIRO, Marsfield, NSW 2122, Australia (e-mail: chandra.thapa@data61.csiro.au).

Digital Object Identifier 10.1109/JIOT.2021.3110256

time [12]. In addition, the detected hot-spot area does not indicate the exact location of a fire. It is also possible that clouds or the smoke prevent the system to have an accurate estimation of the fire location. Thus, limited resolution and the lack of real-time data generation make the satellite imagery approach inefficient for continuous monitoring of a forest zone [12]. Moreover, when the heat source is too small (at the early stages of a bushfire), the system may not consider that zone as a hot spot [11].

Unmanned Aerial Vehicles (UAVs)-Based Solutions: Furthermore, UAVs have been used to take real-time images of fire in the areas that are extremely dangerous to access [12]–[14]. The images are analyzed to assess the situation and make operational decisions to prevent outbreaks of the fire. UAVs can also be equipped with thermal imaging cameras to enable firefighters to see through the smoke and search for the victims. This approach is effective in post-fire situations but has very limited advantage in the early detection of bushfire. In addition, UAVs need high-speed communication links for the transmission of high-quality images that is a big issue in many regional forest areas. Section II covers more detail of the practical solutions being used by Australian fire agencies.

Motivated from this, we propose to use the Internet of Things (IoT) technology and machine learning (ML) models to identify early symptoms of a bushfire. Our proposal is based on spatial correlation theory. We first train the ML model using the Forest Fire real data set collected from a target area over a period of one year [15]. Then, the real-time environmental data (collected from IoT network and tagged with spatiotemporal information) is fed to the trained model for the identification of possible anomalies. Note that the outliers in sensor data can also be generated by nonbushfire events, such as sensor failures or security attacks. In other words, bushfire detection using an approach that is solely based on outliers may result in high false positive rates. To address this issue, we utilize the spatial correlation between sensor measurements and develop a supplementary classification model to ensure that any detected anomaly has been created due to the early symptoms of a bushfire. The proposed model works based on Moran's I index [16], [17] which is an effective tool for the measurement of spatial correlation in distributed architectures.

We implemented a Proof of Concept (PoC) of our approach in Python. We also performed a prototype implementation of the approach in an edge computing setting to ensure its effectiveness to generate fire alerts at the very early stage of a bushfire.

Our contributions in this article are as follows.

1) We propose an ML-based anomaly detection approach for early detection of bushfire. In the proposed approach, the measurement and collection of spatiotemporal environmental data is performed using an IoT network. The collected real-time data is analyzed by the ML model to identify any spatiotemporal pattern that is different than the normal pattern.
2) To reduce the number of false positive detections, a supplementary classification model is proposed to estimate the origin of the detected anomalies, i.e., early symptoms of a bushfire, sensor failures, or a security attack.
3) We implemented the proposed approach in an edge computing setting to furthermore reduce the detection latency.

Our results confirm the accuracy and effectiveness of the proposed approach. We present the related work in Section II. The proposed architecture and implementation details are given in Sections III and IV, respectively. Performance evaluation results and further discussion are given in Sections V and VI, respectively. We summarize this article in Section VII.

II. Related Work

In [13], a comprehensive survey presents the current problems of forest firefighting and state-of-the-art robotic technologies and solutions in firefighting missions. Akhloufi *et al.* [12] reviewed previous research and industry works done in bushfire management using UAV applications. It considers different onboard sensor instruments, fire perception algorithms, and coordination strategies. Below, we highlight the key approaches being used by different Australian authorities in the Bushfire context.

Australian Bureau of Meteorology (BOM) and Geoscience Australia: In Australia, weather forecasts are provided by the BOM. Corresponding fire authorities determine the appropriate fire danger rating (FDR) by considering the predicted weather, including temperature, relative humidity, wind speed, and dryness of vegetation. These FDRs or alerts (via radio, TV, and the Internet) help the government and communities to take essential appropriate actions to minimize the potential impact and loss. Recently, Geoscience Australia has developed the (real-time) *Sentinel bushfire monitoring system* [5], Digital Earth Australia Hotspots, a national bushfire monitoring system. This collects areas generating a high level of infrared radiation (called Hotspots), via satellites at 10 min intervals, to identify possible fire risks [10]. But it is unable to provide real-time information; every hot-spot information is about 17 min old. Furthermore, the hot-spot size does not indicate the exact fire zone. Also in cases where the heat source is too small to detect, the system does not consider that zone as a hot-spot leading to inaccurate hot-spot identification [11].

MyFireWatch Project: This solution is based on an existing Department of Fire and Emergency Services (DFES) program, redeveloped by Landgate and Edith Cowan University (ECU), Australia, for use by the general public [18]. This system gives a view of satellite observed hot-spots by detecting heat sources above a certain temperature level. Furthermore, these hot-spots update at 2–4 h intervals and do not fully indicate the exact level of severity. It does not provide the exact coordinates of potential fire zones, is unable to determine accurately beyond 2 km, hot-spots in the presence of smoke and cloud cannot be detected. Besides the use of satellites, the Australian firefighters also make use of UAVs as a tool for combating fire [14]. The UAVs (particularly drones) provide real-time pictures of fire zones that are extremely dangerous to access.

This approach is effective in post-fire situations to enable firefighters to make early assessments of damage while enroute to emergencies. The cost of UAVs and the requirement of high speed Internet connectivity are key hurdles in many regional areas or underdeveloped countries.

Bushfire Attack Level Toolbox: Researchers also use modeling to analyze the potential impact of fire before it occurs. For example, to measure the impact of fires on buildings, the Bushfire Attack Level Toolbox is developed which provides access to ArcGIS geoprocessing scripts to calculate Bushfire Attack Level ratings [19]. For a comprehensive understanding of why bushfire monitoring systems does not work as intended, we encourage readers to follow [20] and references therein. Also, very recent recommendations provided by the Nature Conservation Society of South Australia (NCSSA) to the federal government to preparedness for, response to, and recovery from bushfire disasters are listed in [21].

FireCloud Project: In FireCloud, a recently done bushfire detection project [22], the existing satellite-based fire detection systems are supplemented with ground-based IoT remote sensing devices to enhance the accuracy and speed of the bushfire detection process. However, it is difficult to judge the effectiveness of this work since its results have not been published (e.g., experimental results). Moreover, it lacks an effective approach to confront the possible incorrect measurements sent by either faulty sensors or illegitimate IoT sensors/attacked sensors (in case of a cyber attack on the deployed IoT network). Indeed, any IoT-based approach for bushfire detection must be equipped with an effective solution to address such threats (atleast publicly accessible forest areas should be protected).

Synthesis of the Existing Solutions: To the best of our understanding, the current existing solutions lack the ability to report the exact location/coordinates of a potential fire area in real-time before the fire occurs. In addition, the post-fire solutions are not cost effective. Furthermore, we emphasize that due to digitization in every domain, the risk of cyber-attack is another factor that might compromise the existing techniques to generate a false alert and false negative predictions. The existing approaches do not integrally consider approaches to distinguish true and false bushfire alerts. In our approach, we utilize a module which enables the network to distinguish legitimate and compromised alerts.

III. Our Proposed Approach

In this section, we present a high level view of our proposed approach for bushfire detection.

Case 1 (No Fire Danger): In the proposed approach, we consider a land/zone where IoT devices or units are deployed taking environmental readings, i.e., temperature and humidity. The sensor streams continuously generated by IoT units are forwarded to a computing device placed at the edge of the network using edge computing technology. The sensor streams collected from the IoT units (of a small forest land/zone) are analysed to detect an outlier (see Fig. 1). In the absence of any outlier detection, the system considers the zone as *no fire danger,* which means no risk of fire is identified. Here, we

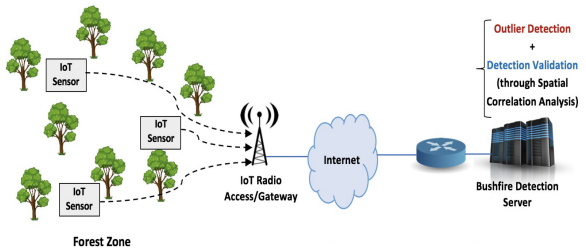

Fig. 1. Proposed system architecture.

have used a natural physical phenomenon of data, i.e., the readings of two nearby IoT units would be the same under normal conditions. For example, humidity and temperature would not change dramatically over short distances. Therefore, we assume that the sensor readings or data would be correlated with its neighbors data, with very high probability. In this case, there is no spatial correlation between the outliers is detected. Contrary if the outlier detection module detects any outlier then the data streams are forwarded to a spatial correlation-based recognition module which is an integral part of the computing module at the edge of the network.

Case 2 (Fire Danger Exists): As mentioned before that in the presence of an outlier, the data is forwarded to a spatial correlation module to accurately identify the potential risk of fire. If the readings are all in the same cluster (no outlier detected at this stage), as expected the framework would still consider this scenario as *no fire danger.* In the case of high fire probability, the level of temperature or humidity will increase naturally which will be measured by sensors and forwarded to the computing module for analyses. In this case, an outlier would be generated. Using this approach of outlier detection, the field or zone would be considered as "*Fire Danger Zone*". However, this case is further divided into three parts.

Case 2.1 (False Fire Danger Zone): The sensors are mostly deployed in harsh environments. It could be possible that there is no fire but the sensor is giving faulty readings. The system could produce outliers and the system may consider this scenario as *Fire Danger Zone*, although it is false. *Counter Argument.* Using the spatial correlation between measurements of neighboring IoT sensors we note that in case of a sensor failure, the failed sensor will generate faulty data, and thus an outlier will be detected but with a very low probability its neighbor sensors also generate faulty measurements as well. Thus, the measurements received from its neighbor sensors are not classified as outliers as no significant spatial correlation between the measurements is detected.

Case 2.2 (Compromised Fire Danger Zone): It could also be possible that the sensors are compromised and giving fake readings. If there are more outliers, then it will be very challenging for the system to differentiate the faulty/false and compromised behavior. *Counter Argument.* We note that in case of a cyber attack on the IoT network, several neighboring sensors are targeted by the attacker. Therefore, neighbor sensors will also generate faulty measurements. In other words, with a very high probability, multiple sensors (located in the neighborhood of the malicious node) will also be affected.

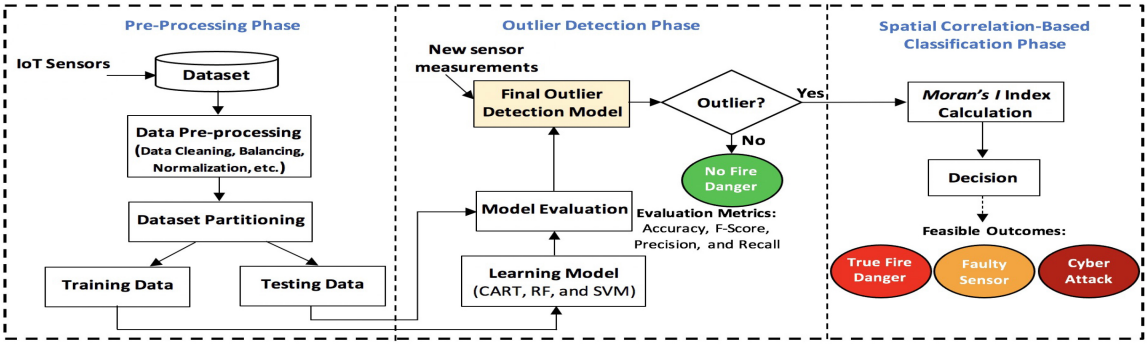

Fig. 2. Work-flow of the proposed approach.

Thus, a significant spatial correlation between the measurements can be observed. In this case, the clustering decisions are made based on the possible spatial correlation between the measurements. In other words, measurements of all the compromised sensors change rapidly and do not follow a location-based and smooth pattern while in case of a true fire, there is a smooth pattern in the change of readings received from neighbor sensors over time, i.e., the outliers are originated from a particular location and gradually propagate to other sensor locations.

Now, the next step is to exactly find the attacked sensors and their location. To effectively exploit the spatial correlation between sensor measurements, we utilize a special mechanism that takes the network address (e.g., IP address) of sensors and returns their x and y coordinates. To do this, we consider a grid that covers the whole geographical area which is under protection. Then, every sensor is assigned with an appropriate x and y coordinate based on its physical location. We emphasize that in the case of mobile IoT sensors this process can be performed using a dynamic method since the location of a sensor may change with respect to time. This can be easily implemented since mobile sensors usually send their real-time location data too along with the data streams.

Case 2.3 (True Fire Danger Zone): It can be argued that all sensors in this case may generate the same readings and thus a high level of spatial correlation will be detected. *Counter Argument.* We first accept that this is true, however, if we increase the land/zone size then it is again highly likely that the readings of one zone from another will be different and the outlier will definitely be detected. We emphasize that the zone size in this case plays an essential role. Therefore, the comparison of different zones or clusters of different zones with each other will also provide more accurate insights of potential risk.

IV. Implementation and Architecture Details

In Fig. 2 we have presented the work-flow of our approach which is divided into three phases (particularly based on spatial correlation between sensor measurements). The first phase is the *preprocessing phase* in which the data streams from IoT sensors are collected and processed. This phase is performed once only during the setup of the IoT network. In the second phase (*outlier detection*), the real-time sensor measurements received from the IoT field network are analysed by an ML model to detect an outlier. In the absence of any outlier detection, the system considers the IoT field network behavior as normal or *No Fire Danger*. Alternatively, if the system detects any outlier, then the data streams are forwarded to a spatial correlation-based recognition module. At this third phase (*spatial correlation-based classification*), the spatial-correlation data of these outliers are fed to a classifier in order to accurately classify the detected outliers into one of the above mentioned clusters, i.e., *False Fire Danger Zone*, *Compromised Fire Danger Zone*, and *True Fire Danger Zone*.

Primarily, we have used the spatial correlation between the detected outlier and the measurements of its neighbor sensors to decide on whether the outlier is due to a *sensor failure* or a *security attack*. We assume that IoT humidity and temperature wireless sensors have been already installed in the field that effectively covers a target area. The sensor measurements are collected by an edge system that is connected to the core system. Regarding the learning model, we have used three ML classification algorithms, i.e., classification and regression trees (CARTs), random forest (RF), and support vector machine (SVM) at the edge segment to detect outliers. In our experiments, we have used the Forest Fire real data set [15]. The data set has been created from the meteorological data gathered in the northeast region of Portugal over a period of four years. In the next section, we present the mathematical modeling of our approach.

A. Mathematical Modelling

As said previously, we utilize the spatial correlation between measurements from neighboring sensors to distinguish between a false fire danger detection and a true bushfire case. To do this, we have utilized the Moran's I index [16], [17] which is an indicator of spatial correlation in distributed architectures. There are global and local versions of the Moran's I index available. The global version indicates the level of spatial correlation over an entire region and is calculated as follows:

$$\mathrm{MI} = \frac{S \sum_{i=1}^{S} \sum_{j=1}^{S} \Delta_{ij} (m_i - \bar{m})(m_j - \bar{m})}{\sum_{i=1}^{S} \sum_{j=1}^{S} \Delta_{ij} \sum_{i=1}^{S} (m_i - \bar{m})^2}$$

TABLE I
DESCRIPTION OF THE FOREST FIRE DATA SET

Feature	Description	Value Range	Mean	Standard Deviation
X	x-Axis spatial coordinate of sensor locations	1–9	4.6	2.31
Y	y-Axis spatial coordinate of sensor locations	1–9	4.3	1.23
Month	Month	Jan–Dec	NA	NA
Day	Day of the week	Mon–Sun	NA	NA
Temp	Temperature (C)	2.2–33.3	18.89	5.81
RH	Relative Humidity (%)	15–100	44.29	16.32
Wind	Wind speed (km/h)	0.4–9.4	4.02	1.8
Rain	Rain (mm/m^2)	0–6.4	0.02	0.3

where S is the number of measuring units (sensors), m_i is the value of the desired feature measured by sensor s_i, and $\bar{m}$ is the mean value of feature m. We also need to determine what sensors are considered as neighbor. This is applied on the Moran's I expression using the Δ_{ij} weights. For example, if s_i and s_j are neighbor, then $\Delta_{ij} = 1$, otherwise, $\Delta_{ij} = 0$. However, we have adopted a more conservative approach in which Δ_{ij} weights are obtained using r_{ij} which is the physical distance between s_i and s_j, i.e.,

$$\Delta_{ij} = \frac{1}{r_{ij}^2}.$$

For the values of MI close to 0, it is concluded that there is no spatial correlation in the spatial data set. On the other hand, if the absolute value of MI is greater than 0, the global correlation of the spatial data set is confirmed.

The expected value of the MI index is calculated as $(-1/S - 1)$ [16], [17]. However, the standardized version of MI is generally used to determine the thresholds on which the spatial correlation becomes significant in a data set. The global Moran's I index is standardized to G(MI) as follows:

$$G(\mathrm{MI}) = \frac{\mathrm{MI} - \mathrm{Exp(MI)}}{\sqrt{\mathrm{Exp}\left(\mathrm{MI}^2\right) - (\mathrm{Exp(MI)})^2}}.$$

For example, a spatial data set is correlated with a confidence level of 95%, if $|G|$ is greater than 1.96. The global Moran's I index reflects the level of spatial correlation over the whole data set. Thus, we use the local version of the index to determine the degree of correlation between neighbor sensors only. For a specific sensor s_i, the local version of the index is calculated as follows:

$$\mathrm{MI}_i = \frac{(m_i - \bar{m})}{p_i^2} \sum_{j=1, j\neq i}^{S} \Delta_{ij}\left(m_j - \bar{m}\right), \quad \text{where}$$

$$p_i^2 = \frac{\sum_{j=1}^{S} \Delta_{ij}}{S-1} - \bar{m}^2.$$

Thus, after an outlier is detected by the outlier detection module for sensor s_i, the local Moran's I index MI_i is calculated for s_i to obtain the degree of spatial correlation between s_i and its neighbors. If the index is close to 0, false fire danger will be reported since in this case, there is no correlation between the measurements of s_i and its neighbors. In other words, it indicates s_i is malfunctioning while other sensors are working well. However, in case of real fire danger, all the neighbor sensors are generally affected by the event. This results in having the measurements of s_i and its neighbors still spatially correlated.

V. PERFORMANCE EVALUATION

In this section, we discuss the test settings, experiments results, PoC implementation, and a comparison of our approach with three related research works.

A. Test Settings

We have developed a PoC using Python 3.7.4. Two types of workstations were used for the experiments: 1) Intel Core i5–11600 2.80-GHz CPU with 8 GB of RAM and 2) Intel Core i7–7700K @4.20-GHz CPU with 64 GB of RAM. We also used an Android mobile device LG G4–H818P equipped with a Hexa–Core 1.8-GHz processor, 3 GB of RAM, and running Android OS 5.1, acting as an IoT device.

Regarding the IoT network, we have used the Eclipse Paho MQTT Python library [23]. The Python code enables the MQTT clients and brokers to publish a message and subscribe to a topic and receive a published message as well. We use the Forest Fire real data set [15] for performance evaluation. The data set has different features, including spatial data (x and y coordinates) and meteorological features, such as temperature and relative humidity (see Table I presents the features of the data set.). For the outlier detection phase, we have used three ML algorithms, i.e., CARTs, RF, and SVM. To model the real fire danger scenario we have added Gaussian noise to the measurements of several randomly selected neighbor sensors.

In addition, we define R as the percentage of sensors that send outlier measurements. We change R from 1% to 5% to see how it affects the accuracy of the proposed solution. Furthermore, to evaluate the intensity of the outliers, we have changed the level of the injected Gaussian noise by considering its standard deviation (sigma) as a percentage of the value of each measurement. For this reason, we have changed sigma from $0.2T$ to $0.5T$ for temperature outliers and from $0.2H$ to $0.5H$ for humidity outliers, where T and H are the true measurements of temperature and humidity, respectively. Naturally, lower levels of noise result in more difficult detection of the relevant outliers.

B. Results

Figs. 3 and 4 show the detection accuracy of the three employed ML algorithms for temperature and humidity parameters, respectively. As we see in the figures, a higher level of

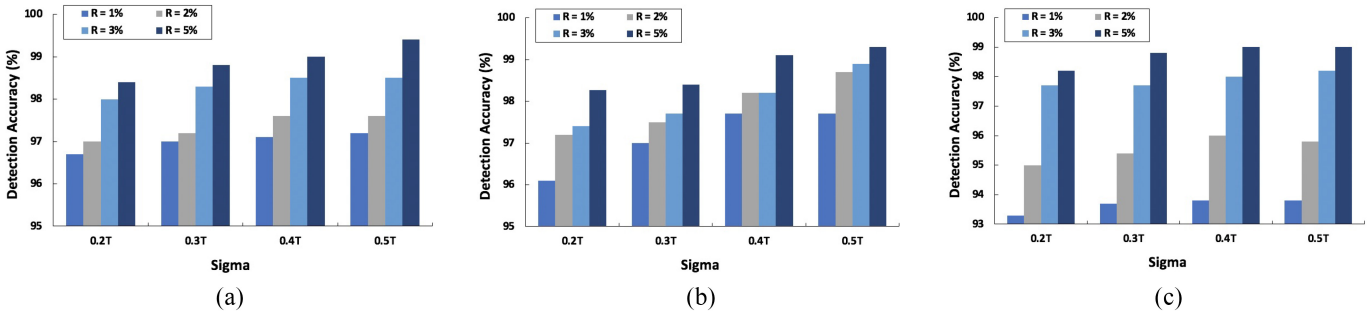

Fig. 3. Detection accuracy based on temperature measurements and the percentage of sensors who generate outlier measurements (a) CART, (b) RF, and (c) SVM algorithms.

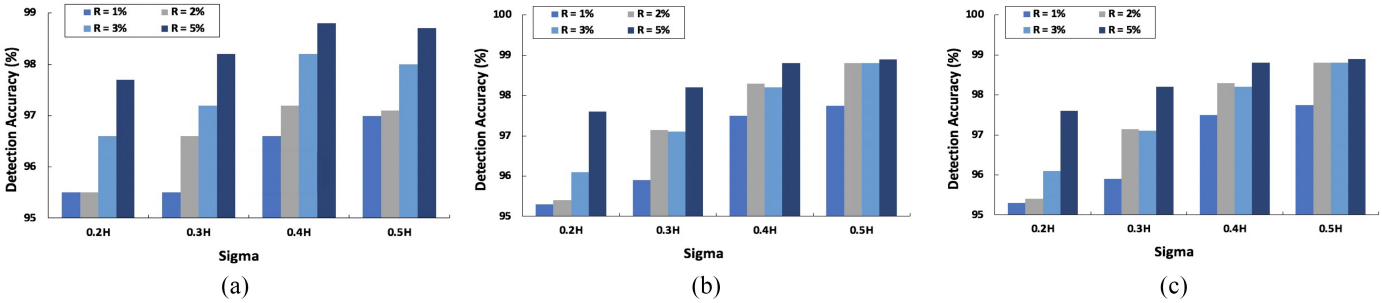

Fig. 4. Detection accuracy based on humidity measurements and the percentage of sensors who generate outlier measurements (a) CART, (b) RF, and (c) SVM algorithms.

the added noise results in higher accuracy, as we expected. Thus, during the beginning of a bushfire and considering a fixed value of R, the detection accuracy increases over time because the new measurements are distancing from the normal expected measurements. This behavior is seen for all the three algorithms. Moreover, having a larger value of R results in higher accuracy. This is because in this case the data set is more skewed, and therefore a new cluster can be created that convinces the classifier algorithms to label it as a separate cluster (i.e., anomaly). In addition, the figures show that the CART and RF algorithms perform better than the SVM algorithm in terms of detection accuracy. They both give similar performance and can detect outliers with the accuracy of as high as 99.4% when 5% of sensors generate outliers. However, the SVM algorithm gives the worst performance among the three algorithms. Specifically, in case of a lower number of sensors that generate outliers (i.e., $R = 1\%$ or $R = 2\%$), SVM performance is poor in terms of detection accuracy. This may result in delayed detections if SVM is employed because it needs a higher number of sensors that generate outliers in order to detect a bushfire case.

We have also performed the PoC implementation of our approach in two different scenarios to evaluate its performance in terms of latency. Fig. 5 shows the network structure used in our prototype implementation. First, we performed the detection process using an edge server that was located close to the IoT device (within the Wifi range). In the second scenario, the procedure was performed by a cloud server with a physical distance of 280 km to the IoT field network. We used *ZeroTier* [24] to connect the IoT field network to the cloud

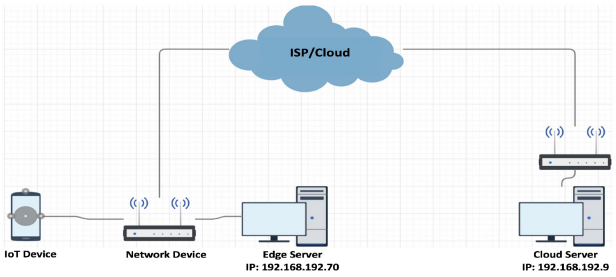

Fig. 5. Network structure of our PoC design.

server. Moreover, *PRTG Network Monitor* [25] was used to monitor the latency between the IoT devices and the servers.

Regarding the latency between the IoT devices and servers, as Fig. 6(a) and (b) show, the proposed approach performs significantly faster in the first scenario in which the detection process is done by the edge server. For example, considering the packet size of 32B, we obtained the average latency of 19 ms in this case while the average latency of 73 ms was obtained in the second scenario in which the sensor measurements are sent to a cloud server by the IoT devices where the detection procedure is performed. Thus, the detection time of a bushfire can be effectively increased by employing several edge servers (each one covers a specific zone in the target area) that collect IoT sensor measurements and perform the proposed detection approach. The detected events, logs, and performance of the edge servers can be monitored in the core segment of the network using a centralized server.

We also recorded the end–to–end latency in the detection of bushfire symptoms. To emulate the behavior of environmental

TABLE II
COMPARISON OF OUR APPROACH WITH THREE RESEARCH WORKS IN THE FIELD OF OUTLIER DETECTION

		Our Approach			[26]			[27]	[28]
		CART	RF	SVM	CART	RF	GBM	EONF ISSNIP	RF
Temperature	Accuracy (%)	98.9	98.8	98.8	97.7	94.7	97.7	98.5	95
	F–Score (%)	98.2	98.2	98.4	95	89	95	Not Available	94
Humidity	Accuracy (%)	98.2	98.4	98.7	99.3	99.3	99.3	98.5	95
	F–Score (%)	97.8	98	98.3	99	99	99	Not Available	94

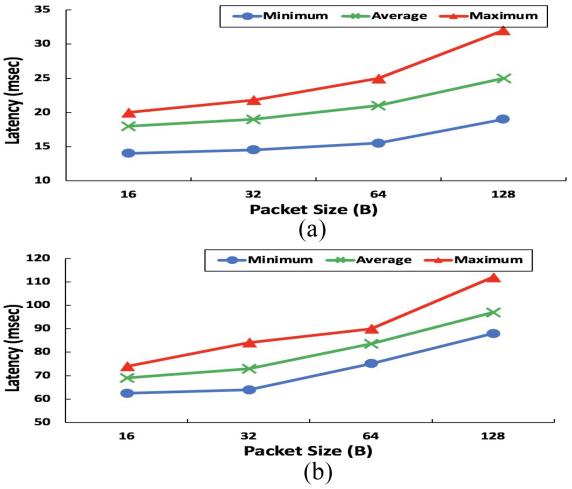

Fig. 6. Latency report between the IoT device and (a) edge computing server and (b) cloud server.

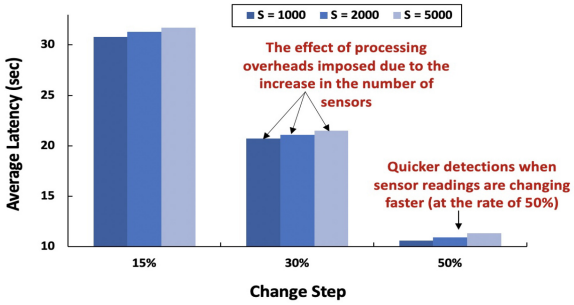

Fig. 7. Average end–to–end latency (change step w.r.t sensor numbers), time interval between sensor readings was 10 s.

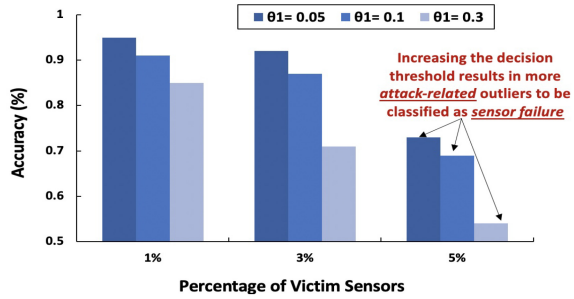

Fig. 8. Average accuracy of attack detection based on the percentage of victim sensors and the Moran's I index threshold θ_1. In the experiments, θ_2 has been set to 5.

parameters at the early stage of a bushfire (e.g., when temperature and relative humidity are increasing and decreasing, respectively), we artificially changed the consecutive sensor readings by steps of 20%, 30%, and 50% (see Fig. 7). As we expected, the lowest detection time was recorded at the change step of 50%. This indicates that the detection time adapts to the spreading speed of fire at its early stage. Moreover, the detection latency heavily depends on the time gap between two successive sensor readings, i.e., refresh time (in our experiments, we used the refresh time of 10 s). In fact, to detect an anomaly, there should be a minimum distance between two consecutive readings. However, this distance may not be created at a single sensor reading (specifically, at low change steps, e.g., 15%). In other words, the ML-based anomaly detection model may need to receive a few numbers of successive sensor readings before it can detect an anomaly. As Fig. 7 shows, the highest level of latency was recorded at the change step of 15% due to the small increase/decrease in the measured values. Furthermore, having a large number of sensors (S) imposes additional processing overhead on the system. This is mainly due to the calculation of MI_is for a large number of sensors.

We performed some experiments to evaluate the system performance in terms of distinguishing between a bushfire case and a security attack (see Fig. 8). In this regard, we used three different values for the number of victim sensors (those who have been successfully compromised by the attacker). This is an important factor since the Moran's I index MI of every outlier depends on the number of victim sensors. In other words, if a limited number of sensors are affected by the attack, the system obtains a small MI for those sensors (close to 0). In this case, the outlier is classified as either a sensor failure (if the obtained MI is lower than the first threshold θ_1) or a security attack (otherwise). However, when a significant number of sensors are affected by the attack, the system obtains a large MI because in this case, a large number of sensors are reporting outliers. In this case, the calculated MI is compared with the second level of threshold to classify the outlier as either a security attack (if the calculated MI is lower than the second threshold θ_2) or a bushfire (otherwise).

C. Comparison

Table II shows the results of a comparison between our approach and three research works in the field of outlier detection [26]–[28]. For the results of our work, we present the average of accuracy and F-score metrics regarding the four results achieved using the four different values of sigma when $R = 1\%$. In [26], an IoT outlier detection architecture is proposed to detect the occurrence of anomalies in a forest environment using four different learning algorithms, i.e., CART, RF, gradient boosting machine (GBM), and linear discriminant

analysis (LDA). We only present the best accuracy achieved in [26] (they have achieved a low accuracy of 78% using the LDA learning model). As seen in Table II, their models have never reached accuracies greater than 97.7% for temperature. For humidity, they have achieved the highest level of 99.3%. However, there is no information and discussion in the paper about the standard deviation (sigma) of the artificial noise added to the data set for outlier generation. In fact, the accuracy will increase if we use a higher level of sigma as confirmed by Figs. 3 and 4.

In [27], the ellipsoidal neighborhood outlier factor (ENOF) mechanism [29] has been used to distinguish normal and anomaly measurements. ENOF is an outlier detection algorithm in which each data point receives an outlier score with respect to the densities of its close neighborhood. Those data points that are located in a dense group of data points receive a small score and vice versa. The ENOF model also needs to calculate a threshold value using the standard deviation of the ENOF scores. This threshold is then used to make anomaly decisions. In their work, ISSNIP [30] and IBRL [31] data sets have been used for the experiments that include both temperature and humidity measurements. Although the ENOF model offers a high level of accuracy, it is still outperformed by our approach (the values of F-score metric have not been presented in their work).

Furthermore, in [28], a modular and hybrid anomaly detection system is proposed for IoT applications. It uses a cloud server for anomaly detection in both application and network layers and train a centralized learning model. The model weights obtained by the cloud server are then transferred to the IoT devices for local anomaly detection. Although it offers efficient performance in terms of detection latency, their achieved accuracies are still lower than our approach. The reason is that there is a deviation between the predictions made by the server and the ones made locally at the IoT devices. Furthermore, it requires the local IoT devices to be frequently updated by the cloud server. This may increase the communication overhead of the system and affects the accuracy.

VI. Further Discussion and Limitations

The shown approach is simple and effective but there are certain aspects that need to be rigorously explored such as: 1) Will the proposed approach truly replace the legacy bushfire detection systems? 2) Do we need to deploy the sensors everywhere in the forests to detect Bushfires? 3) How to deal with two different verticals (tenants) partially using the same IoT deployments but with heterogeneous service requirements? We observe that the first two concerns are interlinked and therefore we have to address them jointly.

First, no solution is best fits in all scenarios therefore we need a holistic approach to understand the emergency of the bushfire problem before we choose the right solution. Nevertheless, we do not argue that our approach will replace the existing solutions, rather we emphasize that it would be beneficial to use the proposed approach on top of the existing approaches in an integrated way. In view of the second concern, we emphasize that further research is required to investigate the modification of IoT-based forest monitoring solutions and identify cost-effective architectural solutions for the effective collection of real-time environmental data from large forest areas. However, a feasible and more cost-effective approach could be to divide the forest area into several smaller zones (i.e., cells). Then, every cell is covered by spreading low-cost sensors which form a single wireless sensor network (WSN) in that cell. In fact, every WSN represents a small-scale cell that can be replicated in a larger scale to cover the entire forest. Each WSN is connected to the main network through one or multiple IoT nodes that serve as wireless gateways. This would be a more economic architectural solution since the deployment of WSNs is much cheaper than deploying a large number of IoT nodes.

We do agree that the above solution would still not be fully practical and cost-effective. Feasibility still remains an open issue. However, we argue that it would be good to identify high fire risk areas first, and then encourage land owners (such as, big farm land, forest zone under forest ministry, etc.), to allow to deploy the given approach, at highly subsidized rates. Residents living nearby potential zones, and emergency fire departments will receive an alert via a locally deployed system, respectively, to get prepare in advance. Gradually deploying the approach from small scale to large scale zone will eventually strengthen the overall effectiveness of the complete bushfire prediction system without replacing the legacy system.

Towards the third concern, it is arguable to say that the deployment of 5G technology and the network slicing concepts within that technology can potentially help to achieve multitenants heterogeneous service requirements [32], [33]. However, we are still far away with a complete roll out of 5G networks. Thus, this is an open challenge, and we need to be cautious with the deployment of the approach using legacy telecommunication providers until 5G reaches in regional areas. We welcome the research community to collaborate with us to address this critical and urgent issue.

VII. Summary and Future Work

For early detection of bushfire dangers in almost real-time, an approach based on ML and spatial correlation between sensor measurements is developed and validated at the edge of the network. The proposed mechanism also ensures that if an attacker successfully attacks the network (e.g., control sensor communications, manipulate their measurements, etc.), the created outliers do not fool the system to generate a false fire alarm. This means the proposal accurately shows that either the fire alarm is due to a true bushfire danger rather than a sensor failure or a cyber attack. The results of our experiments confirmed the merit of our approach. In the future, we intend to investigate the design of novel and cost-effective architectural solutions to ensure the feasibility of IoT-based approaches in the early detection of bushfire, as discussed in the further discussion section.

References

[1] J. Russell-Smith *et al.*, "Bushfires 'down under': Patterns and implications of contemporary australian landscape burning," *Int. J. Wildland Fire*, vol. 16, no. 4, pp. 361–377, 2007.

[2] *Bushfires*. Accessed: Nov. 10, 2020. [Online]. Available: https://www.ga.gov.au/scientific-topics/community-safety/bushfire

[3] *Value of Insurance Losses from Major Weather Events in Australia as of January 2020*. Accessed: Nov. 11, 2020. [Online]. Available: https://www.statista.com/statistics/1098529/australia-insurance-losses-from-natural-disasters/

[4] *2019–20 Australian Bushfires—Frequently Asked Questions: A Quick Guide*. Accessed: Nov. 15, 2020. [Online]. Available: https://parlinfo.aph.gov.au/parlInfo/download/library/prspub/7234762/upload_binary/7234762.pdf

[5] *Real-Time Bushfire Monitoring Satellite System to be Developed by Geoscience Australia*. Accessed: Nov. 11, 2020. [Online]. Available: https://www.abc.net.au/news/2015-08-11/real-time-bushfire-monitoring-system-to-be-developed/6688510

[6] *More than One Billion Animals Killed in Australian Bushfires*. Accessed: Nov. 11, 2020. [Online]. Available: https://www.sydney.edu.au/news-opinion/news/2020/01/08/australian-bushfires-more-than-one-billion-animals-impacted.html

[7] P. Yu, R. Xu, M. J. Abramson, S. Li, and Y. Guo, "Bushfires in Australia: A serious health emergency under climate change," *Lancet Planet. Health*, vol. 4, no. 1, pp. e7–e8, 2020.

[8] J. Alexandra and C. M. Finlayson, "Floods after bushfires: Rapid responses for reducing impacts of sediment, ash, and nutrient slugs," *Aust. J. Water Resourc.*, vol. 24, no. 1, pp. 1–3, 2020.

[9] T. Fukuhara, T. Kouyama, S. Kato, R. Nakamura, Y. Takahashi, and H. Akiyama, "Detection of small wildfire by thermal infrared camera with the uncooled microbolometer array for 50-kg class satellite," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4314–4324, Aug. 2017.

[10] *Bushfire Tracking with Sentinel Hotspots*. Accessed: Nov. 11, 2020. [Online]. Available: https://www.csiro.au/en/Research/Astronomy/Earth-observation/Sentinel-hotspots

[11] *Digital Earth Australia Hotspots*. Accessed: Nov. 15, 2020. [Online]. Available: https://hotspots.dea.ga.gov.au#/

[12] M. A. Akhloufi, A. Couturier, and N. A. Castro, "Unmanned aerial vehicles for wildland fires: Sensing, perception, cooperation and assistance," *Drones*, vol. 5, no. 1, p. 15, 2021. [Online]. Available: https://www.mdpi.com/2504-446X/5/1/15

[13] J. J. Roldán-Gómez, E. González-Gironda, and A. Barrientos, "A survey on robotic technologies for forest firefighting: Applying drone swarms to improve firefighters' efficiency and safety," *Appl. Sci.*, vol. 11, no. 1, p. 363, 2021.

[14] *Drones to Assist Firefighters in Emergencies*. Accessed: Nov. 14, 2020. [Online]. Available: https://www.nsw.gov.au/news/drones-to-assist-firefighters-emergencies

[15] A. Asuncion and D. J. Newman. *UCI Machine Learning Repository. Irvine, University of California, Department of Information and Computer Science*. Accessed: Oct. 16, 2020. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Forest+Fires

[16] M. Yan *et al.*, "Outliers detection of cultivated land quality grade results based on spatial autocorrelation," in *Proc. 5th Int. Conf. Agro-Geoinformat. (Agro-Geoinformatics)*, 2016, pp. 1–5.

[17] M. Das and S. K. Ghosh, "Measuring Moran's i in a cost-efficient manner to describe a land-cover change pattern in large-scale remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 10, no. 6, pp. 2631–2639, Jun. 2017.

[18] *Myfirewatch*. Accessed: Nov. 14, 2020. [Online]. Available: https://myfirewatch.landgate.wa.gov.au/map.html

[19] T. Yang and W. C. Arthur. (2015). *Bushfire Attack Level Toolbox*. Accessed: Nov. 14, 2020. [Online]. Available: https://ecat.ga.gov.au/geonetwork/srv/eng/catalog.search#/metadata/83736

[20] J. Whittaker, M. Taylor, and C. Bearman, "Why don't bushfire warnings work as intended? Responses to official warnings during bushfires in new south wales, australia," *Int. J. Disaster Risk Reduction*, vol. 45, May 2020, Art. no. 101476.

[21] N. de Preu. (2020). *Submission to the Bushfires Royal Commission*. [Online]. Available: https://naturaldisaster.royalcommission.gov.au/submissions

[22] *Firecloud Project*. Accessed: Feb. 1, 2021. [Online]. Available: https://hackerspace.govhack.org/projects/firecloud

[23] *Eclipse Foundation*. Accessed: Feb. 18, 2021. [Online]. Available: http://www.eclipse.org/paho

[24] *Zerotier: Securly Connect Any Device, Anywhere*. Accessed: Sep. 21, 2020. [Online]. Available: https://www.zerotier.com/download/

[25] *PRTG Network Monitor*. Accessed: Sep. 21, 2020. [Online]. Available: https://www.paessler.com

[26] N. Nesa, T. Ghosh, and I. Banerjee, "Outlier detection in sensed data using statistical learning models for IoT," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2018, pp. 1–6.

[27] L. Lyu, J. Jin, S. Rajasegarar, X. He, and M. Palaniswami, "Fog-empowered anomaly detection in IoT using hyperellipsoidal clustering," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1174–1184, Oct. 2017.

[28] A. Ayad, A. Zamani, A. Schmeink, and G. Dartmann, "Design and implementation of a hybrid anomaly detection system for IoT," in *Proc. 6th Int. Conf. Internet Things Syst. Manage. Security (IOTSMS)*, 2019, pp. 1–6.

[29] S.Rajasegarar *et al.* "Ellipsoidal neighbourhood outlier factor for distributed anomaly detection in resource constrained networks," *Pattern Recognit.*, vol. 47, no. 9, pp. 2867–2879, 2014.

[30] V. K. Sachan, S. A. Imam, and M. T. Beg, "Energy-efficient communication methods in wireless sensor networks: A critical review," *Int. J. Comput. Appl.*, vol. 39, no. 17, pp. 35–48, 2012.

[31] IBRLDataset. (2020). *City of Melbourne Open Data Homepage*. [Online]. Available: https://data.melbourne.vic.gov.au/Environment/Sensor-readings-with-temperature-light-humidity-ev/ez6b-syvw

[32] G. Fortino *et al.*, "Towards multi-layer interoperability of heterogeneous IoT platforms: The INTER-IoT approach," in *Integration, Interconnection, and Interoperability of IoT Systems*. Cham, Switzerland: Springer, 2018, pp. 199–232.

[33] L. Zanzi, F. Giust, and V. Sciancalepore, "M^2 EC: A multi-tenant resource orchestration in multi-access edge computing systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2018, pp. 1–6.

Mohammad Reza Nosouhi (Member, IEEE) received the master's degree from Isfahan University of Technology, Isfahan, Iran, in 2006, and the Ph.D. degree from the University of Technology Sydney, Ultimo, NSW, Australia, in 2020.

He is currently working as a Research Fellow with Keshav Sood with the Centre for Cyber Security Research and Innovation, Deakin University, Geelong, VIC, Australia.

Keshav Sood received the B.Tech. degree in electronics engineering (Distinction) and the M.Tech. degree in optical fiber engineering from Punjab Technical University, Jalandhar, India, in 2007 and 2012, respectively, and the Ph.D. degree in information technology from Deakin University, Geelong, VIC, Australia, in 2018.

He worked as a Research Fellow with the Advanced Cyber Security Engineering Research Centre, The University of Newcastle, Callaghan, NSW, Australia. He is currently a Lecturer with the Centre for Cyber Security Research and Innovation, School of IT, Deakin University.

Neeraj Kumar (Senior Member, IEEE) received the Ph.D. degree in CSE from Shri Mata Vaishno Devi University, Katra, India, in 2009.

He is currently a Full Professor with the Department of Computer Science and Engineering, Thapar University, Patiala, India. He is also with the Department of Computer Science and Information Engineering, Asia University, Taichung City, Taiwan, and with the School of Computer Science, University of Petroleum and Energy Studies, Dehradun, India.

Tricia Wevill received the Bachelor of Science (Hons.) and Ph.D. degrees from Monash University, Melbourne, VIC, Australia, in 1999 and 2009, respectively.

She is a Lecturer of Environmental Sciences with Deakin University, Geelong, VIC, Australia. Her research interests include the study of vegetation dynamics at both fine-scale and landscape levels, in response to altered environmental and disturbance regimes. She is particularly interested in how ecological theory can be used to understand community development and the application of this knowledge to land management.

Chandra Thapa (Member, IEEE) received the Ph.D. degree from the University of Newcastle, Callaghan, NSW, Australia, in 2018.

He is a Postdoctoral Fellow with the Distributed Systems Security, Data61, CSIRO, Marsfield, NSW, Australia. His working domain is privacy-preserving distributed AI/machine learning, security, and trust.