Towards physical intrusion detection method based on machine learning and context-aware activity recognition in real-time

NENAD KATANIĆ, KREŠIMIR FERTALJ
Department of Applied Computing, Faculty of Electrical Engineering and Computing
University of Zagreb
Unska 3, 10000, Zagreb
CROATIA
nenad.katanic@fer.hr http://www.fer.unizg.hr/nenad.katanic

Abstract: Sensor-based human activity recognition is getting increasingly popular in various applications. Most of the related work within dense-sensing based approaches assume that large number of different multimodal sensors are placed on the objects in the environment (which is rarely the case in today's real life home environments), that sensor data is not processed in real-time and that activity to be classified is always performed within the same context, thus perform poorly when tested in real life scenarios. In this paper we report on the current status and future steps towards a generic context-aware method for human activity recognition, based on a real-time raw sensor data stream coming from a minimum number of sensors placed in the environment. We propose a hybrid method based on state-of-the-art data-driven and knowledge-driven approaches. Proposed method is being developed and will be validated on the example of the application for robust physical intrusion detection on home doors in real life environment.

Key-Words: activity recognition, machine learning, context-aware, real-time, dense-sensing, accelerometer, physical intrusion

1 Introduction

Physical activity recognition is becoming increasingly popular in various domains such as ambient assistive living [1] [2], context-aware computing [3], mobile computing [4], and others [5]. As authors explain in [6], maturity of low-cost sensor technologies and wireless communication networks has pushed the research focus to high-level information integration, context processing, and activity recognition and inference.

In this work we aim at developing a generic context-aware method for human activity recognition in real-time, based on a real-time raw sensor data stream coming from a minimum number of sensors placed in the environment activity is performed. In terms of method validation, we will investigate the possibility of applying such method for efficient real-time physical intrusion detection and physical intrusion attempt type classification (e.g. burglary attempts using a lock-picking [7] or bump-key [8] methods). Since we aim at using the minimum possible number of low-cost sensors for data gathering, we believe that such method can open a possibility for a whole range of other different widely accessible (low-cost) systems and applications, physical intrusion detection system being just one of them. In order to further open up a possibility for wider use and accessibility of applications built upon our method, we aim at performing real-time activity recognition and classification entirely on the server side based on raw sensor data, without any additional data pre-processing on sensor nodes themselves. In this scenario, any application based on this method becomes independent of the hardware, simultaneously making the hardware architecture simple, low-cost and therefore widely accessible.

The remainder of the paper is structured as follows. Section 2 gives a short introduction into the subject, including explanation of the terminology and classification of the approaches in the field of sensor-based activity recognition. Here we also reflect on the related work and identify drawbacks of existing works in the field. In section 3 we report on our proposed approach and research methodology for achieving our goal. Finally, we close the paper with few concluding words in Section 4.

2 Problem Formulation

2.1 Sensor-based activity recognition approaches

Sensor-based activity recognition is a human activity recognition approach based on the use of

emerging sensor network technologies for activity monitoring, where generated sensor data are primarily time series of state changes and/or various parameters that can be further processed in a desired manner [6].

With respect to the position of the sensor(s) we can differentiate between two main approaches [6]:

1. *Wearable sensors* – where sensors are attached to an actor under observation;
2. *Dense-sensing* – where sensors are attached on objects that constitute the environment in which an activity is being performed. Generally, in this approach we infer on performed activities by monitoring human-object interactions through large number of low-cost multimodal sensors.

In our approach, sensor(s) will always be placed on the object constituting the activity environment, which therefore falls into the category of dense-sensing based approaches.

Regarding the method for building activity models used to infer on performed activities, we also differentiate between two main approaches [6]:

1. *Data-driven* – learning activity models based on a general dataset of recorded activities, using data mining and machine learning techniques, after which activity inference can be performed using probabilistic or statistical classification. Since this approach relies on preexistent dataset, it is important to point out that therefore this approach suffers from "cold start" problem (data scarcity). Additionally, it suffers from reusability problems in terms of applying activity models between different actors.
2. *Knowledge driven* – building activity models based on rich prior knowledge in the domain of interest through formal modeling and representation, where activity recognition is usually performed through formal logical reasoning. Unlike the data-driven approaches, this approach has not proven well in handling uncertainty and temporal information.

As we will further explain in the remainder of the paper, regarding the method for building activity models, we aim at the hybrid approach, based on the combination of data-driven and knowledge-driven approaches.

2.2 Generic human activity recognition process

In general, activity recognition process consists of four main tasks [6]:

1. Choosing and deploying appropriate sensors to actors and/or objects in the environment in order to capture user's behaviour and/or state changes in the environment
2. Collecting, storing and processing sensor data through data analysis techniques and/or knowledge representation formalisms
3. Creating computational activity models in a way which enables software systems to conduct reasoning
4. Select or develop reasoning algorithms to infer activities from sensor data.

2.3 Related work and identified problems

In recent years, many different approaches to sensor-based human activity recognition have been investigated, showing very promising results. Even though in our work we focus primarily on dense-sensing based approaches, it is also worth pointing out some research results based on wearable-sensors approach since these two are complementary in a way that they share the same generic human activity recognition process as outlined in section 2.2, but differentiate in terms of sensor positions.

Authors in [9] have successfully performed human activity recognition for twenty different activities based on data collected from multiple biaxial accelerometers placed on different body parts of subjects under consideration. Four different classifiers have been tested (KNN, Naive Bayes, decision tree and decision table). All classifiers were built based on four manually extracted features (mean, spectral energy, frequency-domain entropy and correlation of accelerometer data). In this case, decision tree showed best results. Lowering the number of sensors used, authors in [10] collected the data from a single tri-axial accelerometer worn near the pelvic region, also using four different features but with the difference of using standard deviation instead of frequency domain entropy. They tested eighteen different classifiers in four different settings. In this approach, Plurality Voting showed as best choice in some settings, while Boosted SVM performed better in others.

When it comes to dense-sensing based approaches, they are mostly employed in Ambient Assisted Living (AAL) applications, that is in the development of smart home environments. Examples of such efforts include [11] where authors have successfully used various sensors in the environment (such as motion detectors, binary switches and pressure mats) for activity recognition. On the other hand, authors in [12] are using large number of infrared sensors to recognize activities such as loitering, turning and walking.

There is a large number of other efforts based on usage of various other types of sensors [13], using different number of the same type of sensor [14], [15], and also using different machine learning methods and their combinations [16] [17].

It is very important to point out that with all these different combinations and approaches to human activity recognition, it is impossible to claim that one sensor deployment for a specific application scenario is superior to the other. Instead, suitability and performance depend on the nature and type of the activities being assessed and on the characteristics of particular application/domain [6].

Most of the current work is assuming that learning data sets and test data sets have both been gathered within the same context and in most cases recorded in laboratory environment, thus perform poorly when tested in real life scenarios and when activity to be recognized is performed by a person that is not the one who performed activities based on which the activity model has been built in the first place. Additional drawback is that most developed methods do not support activity recognition in real-time, but instead require all sensor data which represents a specific activity to be recorded before activity recognition can take place. Furthermore, when it comes to context modeling, we identified that in some cases different authors consider different concepts as context. For authors in [18] context consists of information contained in neighboring data frames regarding the frame on which activity recognition is performed. They showed that in comparison to the previous works where input to the classifier were only the features extracted from the current frame, when we also incorporate information of the neighboring frames activity recognition performance is improved and the more the neighboring frames were taken into consideration, the higher the recognition accuracy was. On the other hand, authors in [19] employ knowledge-driven approach and use ontologies for explicit context and activity modeling and representation. They model and differentiate between *spatial contexts* (such as location information and surrounding entities like household furniture and appliances), *event contexts* (containing background activities and dynamic state changes of appliances and devices), *environmental contexts* (composed of environmental information such as temperature, humidity, etc.) and *temporal contexts* (which indicate time/or duration). When incorporating such contextual information in activity recognition process, they showed that their approach performs well both in laboratory environments as well as in real life scenarios.

For additional information and more comprehensive overview of the related work regarding different approaches to sensor based human activity recognition please refer to [6].

3 Problem Solution

3.1 Proposed approach and expected contribution

Based on related work outlined in section 2.3 and identified drawbacks of existing approaches to dense-sensing based activity recognition, we aim at developing a *context-aware* method for human activity recognition, based on a *real-time raw data stream* gathered from a *minimum possible number of sensors* placed *in the environment* in which activity is performed. As previously outlined that suitability and selection of different sensors and methods for activity recognition depend on the nature and type of the activities assessed and on the characteristics of particular application/domain, proposed method will be developed and validated regarding selected application domain of physical intrusion detection on home doors. Since most of other works based on dense-sensing based approach assume large number of different multimodal sensors placed on the objects in the environment (which is rarely the case in today's real life home environments) while we aim at the opposite, we call our approach a *sparse-sensing* approach. Expected final scientific contributions of our work are the following:

1. Machine-learning based methodology for feature extraction from sensor data, suitable for context-aware physical intrusion detection in real-time.
2. Context model suitable for robust physical intrusion detection in real environment based on sensor data.
3. Methodology for data collection as well as for building and verifying the machine learning model for context-aware physical intrusion detection.

3.2 Data collection methodology

In order to collect required data sets both for activity model development and validation of the proposed method, we built a small wireless sensor network in-house based on the hardware we have at our disposal. We developed four different autonomous wireless sensor nodes, each consisting of the following components:

- Raspberry Pi 3 device [20]
- ADATA PV120 Rechargeable Li-polymer Battery Power Bank with 5100mAh capacity [21] for power supply, making the sensor node autonomous and portable
- Single tri-axial ADXL345 accelerometer from Analog Devices [22] (placed on the extension board provided by AdaFruit Industries [23]) connected to Raspberry Pi's GPIO input pins. This accelerometer supports wide sensitivity range (±2G - ±16G), wide output data range (0.1Hz – 3200Hz) and high resolution measurement (13 bit) which makes it suitable for various different purposes and applications.

Each sensor node is attached on the inside of a different door within our faculty department and connected to faculty's Wi-Fi network, with the accelerometer placed near the door lock. Each node transmits raw sensor data to a single desktop PC through direct TCP connection. Predefined initial set of activities is performed by different employees (actors) of the department. In order to correctly label the data and distinguish between different activities, each recording session is performed according to the following predefined steps with the delay between consecutive activities of at least five seconds (but without explicitly defined way activities shall be performed):

Precondition: doors are *closed* and *unlocked*

1. Actor makes a note of his identity and the current time shown by the sensor node when the recording session starts.
2. Actor performs following set of activities in this particular order:
    a. *knocking*: Knock on the door
    b. *incorrect-key-insertion*: Insert incorrect key into the door lock
    c. *incorrect-key-removal*: Remove incorrect key from the door lock
    d. *correct-key-insertion*: Insert correct key into the door lock
    e. *lock*: Lock the door
    f. *unlock:* Unlock the door
    g. *open-door:* Open the door
    h. *close-door:* Close the door
3. Actor makes a note of the current time shown by the sensor node when the recording session ends.

On each door (sensor node) recording session will take place at least once per working day. This means that we will be able to gather 4 (sensors) x 5 (days) = 20 samples per activity each week. Over time this will expectedly become a dataset large enough to both learn and test the classifiers, with the data recorded in real world environment. Some preliminary results of data gathering are show on the following figures (Figures 1-5).

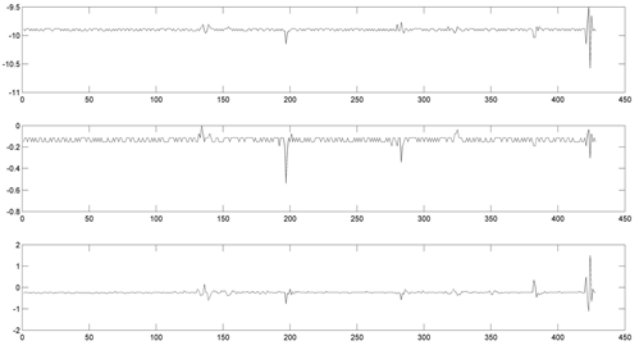

Fig. 1. Accelerometer data for *correct-key-insertion* activity

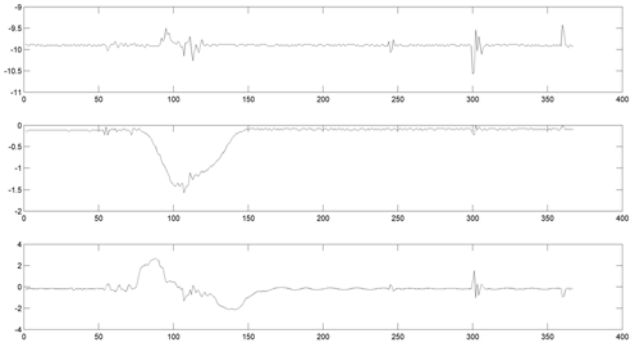

Fig. 2. Accelerometer data for *open-door* activity

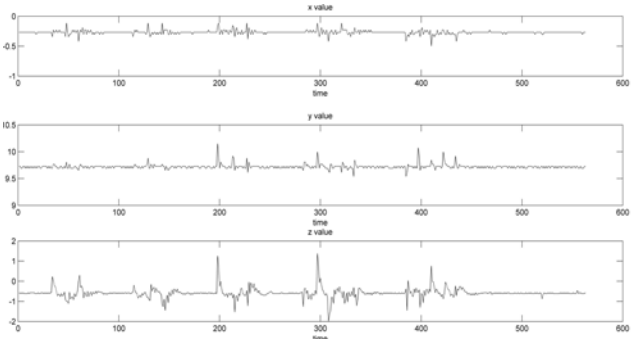

Fig. 3. Accelerometer data for *knocking* activity

As we can see from the figures above (Fig. 1, Fig. 2, Fig. 3) recorded data intuitively shows that we will be able to distinguish between these different activities. This intuition especially holds for the pair of activities like *correct-key-insertion* (Fig. 1) and *open-door* (Fig. 2) which are easily distinguishable even with the naked eye. In combination with these two activities, we can assume the same for the knocking activity (Fig. 3) as well. On the other hand, we intuitively expect that recorded data for

incorrect-key-insertion activity will look very similar to the *correct-key-insertion* (Fig. 1) which will thus be much harder to recognize and distinguish.

Once we achieve satisfactory results on the above mentioned activities, we plan to include and record data for additional set of more complex activities more suitable for the targeted domain of physical intrusion detection, performed by the professional locksmith. These might include activities such as burglary attempt using a lock-picking method [7] and burglary attempt using a bump-key method [8]. Preliminary examples of recorded data for these activities (performed by the author itself) are shown on the following figures. Fig. 4 shows that *lock-picking* activity contains a lot of noise and uncertainty and will thus expectedly be somewhat harder to classify, but still, we intuitively also see that it looks quite different from all other activities. Thus, we assume that even such activities can be recognized based on the raw data gathered from a single tri-axial accelerometer. Same holds for the *bump-key* activity (Fig. 5).

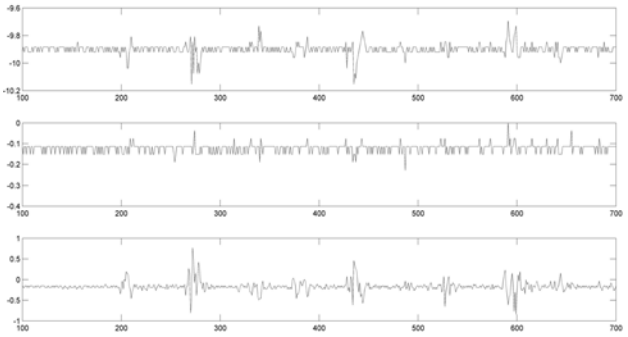

Fig. 4. Accelerometer data for *lock-picking* activity

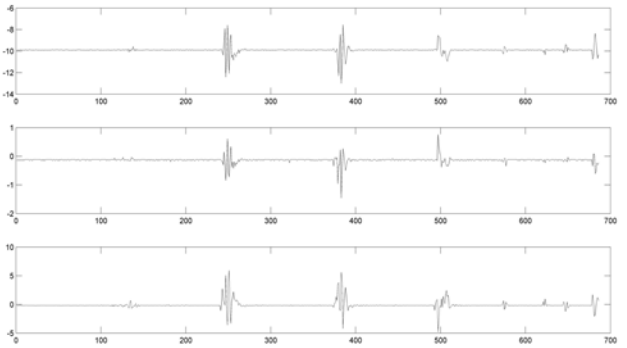

Fig. 5. Accelerometer data for *bump-key* activity

3.3 Research methodology

After the datasets have been collected, we will continue towards our final goal of developing a robust real-time method for physical intrusion detection based on machine learning and context-aware activity recognition, as follows:

3.3.1 Feature extraction

As already explained in the introduction section, we aim at performing real-time activity recognition and classification entirely on the server side based on raw accelerometer data, without any additional data preprocessing on sensor nodes themselves. Therefore, our main focus for feature extraction will be on the Deep Neural Networks [24], a powerful tool for feature representation, which can learn features and build a classifier by itself based solely on raw sensor data. This approach has already proven to be successful both for feature extraction and activity recognition based on raw sensor data, as shown by authors in [18].

In case of non-satisfactory results with the above mentioned approach based on Deep Neural Networks, we will also investigate the possibility of manual feature extraction. In this case, as features to be selected depend on the targeted application domain [6] and in order to find optimal set of features for selected domain, we will investigate and perform experiments with selected features from both time and frequency signal (accelerometer data) domains, extracted from the raw sensor data gathered through data collection methodology explained in section 3.2.

3.3.2 Context modeling

As shown in the related work section, different authors consider different concepts as context, with two selected works being the most representative. In our work we will investigate the possibility of the hybrid approach. We will build upon the work of authors in [18] where context consists of information contained in neighboring data frames with respect to the frame on which activity recognition is performed since they proved that this approach improves activity recognition results. On the other hand, we aim not only at high activity recognition accuracy, but also want to achieve a robust method that can perform well in real life scenarios. In this respect, we will combine above mentioned approach with the knowledge-driven approach taken by authors in [19]. We will investigate the possibility of using ontologies representing Spatial, Event, Environmental and Temporal contexts (with identified properties regarding the selected domain of physical intrusion detection) and test if such hybrid approach can enable robust, high accuracy activity recognition in real life environment.

3.3.3 Activity recognition and validation

With the data sets and context models in place we will continue with building and testing the classifiers for activity recognition. For testing and validation purposes, subset of the collected data set will be selected as a training data set, while the

remaining subset will be selected as the test data set. In ideal case, Deep Neural Networks that we aim to use for feature extraction and building the classifiers will give satisfactory results when it comes to activity recognition/classification. If not, we will also perform benchmarking of different other machine-learning methods and algorithms previously proved to be successful for human activity recognition. This might include methods based on generative modeling (such as Naïve Bayes) as well as discriminative modeling approaches (such as Nearest Neighbor, Decision Tree and SVM). For this purpose, some of publicly available machine learning toolkits will be used, such as WEKA machine learning toolkit [25] [26] previously used in various works in the field of human activity recognition, such as [18].

4 Conclusion

Even though numerous existing approaches to sensor-based human activity recognition exist, most of the existing work suffers from at least one of the following drawbacks. They are often based assuming that both the learning and test data sets have been gathered within the same context (that is, they do not take contextual information into account) and in most cases these data sets are recorded in laboratory environment, thus the resulting classifiers perform poorly when tested in real life scenarios (e.g. when activity to be recognized is performed by different person that the one who performed activities on which the activity model has been built in the first place). Additional drawback is that most developed methods do not support activity recognition in real-time, but instead require all sensor data which represents a specific activity to be recorded before activity recognition can take place. Finally, when it comes to works based on dense-sensing based approaches, they usually assume that a large number of different multimodal sensors are placed within the environment, which is rarely the case in today's real life home environments.

In our approach we aim at developing a context-aware method for human activity recognition, based on a real-time raw data stream gathered from a minimum possible number of sensors placed in the environment in which activity is performed. We believe that such method can open a possibility for a whole range of other different widely accessible (low-cost) systems and applications. Since we will also incorporate contextual information, we expect that applications built on top of such method will be robust enough to be used in real world scenarios. We will showcase the applicability of such method on the example application for robust physical intrusion detection on home doors.

References:

[1] M. Philipose, K. P. Fishkin, M. Perkowitz, D. J. Patterson, D. Fox, H. Kautz, and D. Hahnel, Inferring activities from interactions with objects, *IEEE Pervasive Comput.*, vol. 3, no. 4, pp. 50–57, Oct./Dec. 2004.

[2] T. Van Kasteren and B.Krose, Bayesian activity recognition in residence for elders, in *Proc. Int. Conf. Intell. Environ*., Feb. 2008, pp. 209–212.

[3] K. Van Laerhoven and K. A. Aidoo, Teaching context to applications, *J. Pers. Ubiquitous Comput.*, vol. 5, no. 1, pp. 46–49, 2001.

[4] T. Choudhury, S. Consolvo, and B. Harrison, The mobile sensing platform: An embedded activity recognition system, *IEEE Pervasive Comput*., vol. 7, no. 2, pp. 32–41, Apr./Jun. 2008.

[5] R. Poppe, A survey on vision-based human action recognition, *Image Vis. Comput*., vol. 28, no. 6, pp. 976–990, 2010.

[6] L. Chen, J. Hoey, C. D. Nuget, D. J. Cook, Z. Yu, Sensor-based Activity Recognition, *IEEE Transactions on Systens, Man, and Cypernetics – Part C: Applications and Reviews*, vol. 42, No. 6, November 2012.

[7] Lock picking, https://en.wikipedia.org/wiki/Lock_picking, Nov. 2016.

[8] Lock bumping, https://en.wikipedia.org/wiki/Lock_bumping, Nov. 2016.

[9] L. Bao and S. S. Intille, Activity recognition from user-annotated acceleration data, *Pervasive Computing, Lecture Notes in Computer Science*, vol. 3001, pp. 1–17, 2004.

[10] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, Activity Recognition from Accelerometer Data, *Proceedings of the Seventeenth Conference on Innovative Applications of Artificial Intelligence*, vol. 5, pp. 1541-1546, 2005.

[11] E. M. Tapia, S. S. Intille, and K. Larson, Activity recognition in the home using simple and ubiquitous sensors, in *Proc. Pervasive*, 2004., pp. 158–175.

[12] C. R. Wren and E. M. Tapia, Toward scalable activity recognition for sensor networks, in *Proc. 2nd Int. Workshop Location Context-Awareness*, 2006, pp. 168–185.

[13] Yin X., Shen W., Sarnarabandu J., Wang X., Human Activity Detection Based on Multiple Smart Phone Sensors and Machine Learning Algorithms, *Proceedings of the 2015 IEEE 19th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 2015.

[14] Olutoyin Oshin T., Poslad S. Energy-Efficient Real-Time Human Mobitlity State Classification Using Smartphones, *IEEE Transactions on computers*, vol. 64, no. 6, June 2015.

[15] Trabelsi D., Mohammed S., Chamroukhi F., Oukhellou L., Amirat Y., An Unsupervised Approach for Automatic Activity Recognition Based on Hidden Markov Model Regression, *IEEE Transactions on automation science and engineering*, vol. 10, no. 3, July 2013.

[16] Cenedese A., Antonio Susto G., Belgioioso G., Ilario Cirillo G., Fraccaroli F., Home Automation Oriented Gesture Classification From Inertial Measurements, *IEEE Transactions on automation science and engineering*, vol. 12, no. 4, October 2015.

[17] Vepakomma P., Debraj D., Sajal K., Bhansali S., A-Wristocracy: Deep Learning on Wrist-worn Sensing for Recognition of User Complex Activities

[18] Licheng Z,. Xihong W., Dingsheng L., Improving Activity Recognition with Contextual Information, *Proceedings of 2015 IEEE International Conference on Mechatronics and Automation*, August 2 - 5, Beijing, China

[19] L. Chen, C. D. Nugent, and H. Wang, A knowledge-driven approach to activity recognition in smart homes, *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 6, pp. 961–974, Jun. 2012.

[20] Raspberry Pi Documentation, https://www.raspberrypi.org/documentation/, Nov. 2016.

[21] ADATA PV120 Rechargeable Li-polymer Battery Power Bank Specification, http://www.adata.com/us/mobile/specification/324, Nov. 2016.

[22] Analog Devices ADXL345 Accelerometer Specification, http://www.analog.com/media/en/technical-documentation/data-sheets/ADXL345.pdf, Nov. 2016.

[23] AdaFruit ADXL345 Digital Accelerometer Documentation, https://learn.adafruit.com/adxl345-digital-accelerometer, Nov. 2016.

[24] Neural Networks and Deep Learning, http://neuralnetworksanddeeplearning.com/

[25] S. R. Garner, Weka: The waikato environment for knowledge analysis, *Proceedings of the New Zealand computer science research students conference*, pp. 57-64, 1995.

[26] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, The WEKA data mining software: an update, *ACM SIGKDD explorations newsletter*, vol. 11, pp. 10-18, 2009.