An Immune Inspired Algorithm for Fault Tolerant Enhanced Multimodal Machine Learning

Mattias Cross
Department of Computer Science
University of Sheffield
mgcross1@sheffield.ac.uk

Valentin Radu
Department of Computer Science
University of Sheffield
v.radu@sheffield.ac.uk

***Abstract*—Much like how sentient beings use multiple senses, *Multimodal Machine Learning* (MML) uses multiple input signals to form an estimation about the environment. An advantage MML has over standard Unimodal Learning is that if one input becomes corrupt or unavailable, a *multimodal* recognition model can rely on the other inputs by *zeroing out* the corrupt or missing one. As such, MML can become more robust and fault-tolerant. Unfortunately, it is not always known when an input has become corrupt or not. Corrupt inputs are anomalous in comparison to normal and expected data. We use an immune inspired algorithm, the *Negative Selection Algorithm* (NSA) as a resilient anomaly detection algorithm. Using the multimodal *Smart Gym* "MM-FIT" dataset, we evaluate how corrupt inputs affect the accuracy of MML classification. We implement the NSA specifically for a multimodal pipeline of activity recognition with anomaly detection. The results indicate that our recognition pipeline is effective in introducing fault-tolerance to MML.**

***Index Terms*—Multimodal Machine Learning, Negative Selection Algorithm, Fault Tolerance, Anomaly Detection, Human Activity Recognition**

I. Introduction

Sensor data is prone to errors due to irregular behaviour of the hardware, interference and data acquisition glitches. These unexpected events alter the signal in a small or more substantial proportion, which may have severe consequences for critical applications. We refer to these events as anomalies of the sensor signal.

Tackling anomalies of sensor data is not simple. Chandola et al. [1] define anomalies as "patterns in data that do not conform to what is expected". We see anomalies as transformations of data that could lead to mis-classifications, or having a detrimental effect on the training in Multimodal Machine Learning (MML). MML uses multiple *modalities* (sensors) to solve a task, e.g. activity recognition. Any system collecting data from multiple senses offers the opportunity to incorporate fault-tolerance. If one or more sensor signals are affected by noise, or other corruptions, the rest of the stable sensors may still hold enough information to conclude the recognition task without being affected by the corrupted ones.

To investigate fault-tolerance in MML via anomaly detection, this paper uses the MM-FIT *Smart Gym* dataset [18]. We implement an immune inspired Negative Selection Algorithm (NSA). This has been previously explored by Forrest et al. [4] and Dasgupta et al. [3]. We use the NSA algorithm for detecting corrupted signals across sensing modalities and then

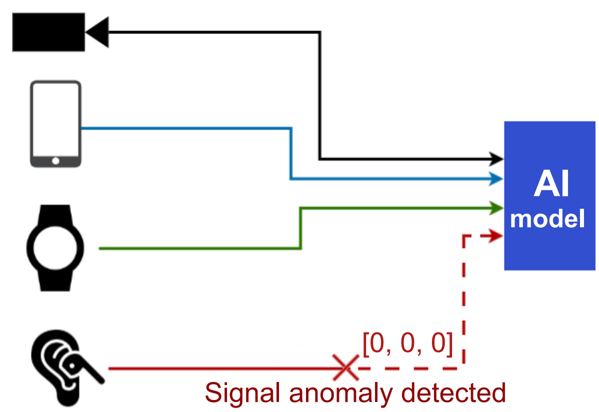

Fig. 1. The recognition system based on Multimodal Neural Networks detects the anomalous data inputs first and then it "zeros-out" the faulty inputs before they can corrupt the recognition task. This example shows a possible situation where the sensing modality from the earbud is rejected because of anomaly detection while the remaining modalities (camera, phone and watch) continue to provide inputs for the activity recognition task.

zero them out if deemed anomalous, to enhance fault-tolerance and increase the robustness of our recognition system.

MM-FIT [18] is a multimodal dataset, capturing inertial signals from different wearable devices during physical exercise activity. These devices are time-synchronised and capture the entire body movement due to their different locations on the human body. The sensor signals are fed into multiple *autoencoders* and a *Convolutional Neural Network* [14] to classify the type of exercise (e.g. squats), the user is performing during the workout. The baseline is produced with a multimodal deep neural network [15]. Besides activity recognition, this dataset offers annotations for counting the number of repetitions of an exercise, while also classifying the rest periods as non-exercise. This task and dataset provide the best testbed for our proposed algorithm of anomaly detection in multimodal sensor signals.

The Negative Selection Algorithm (NSA) is inspired by the negative selection of *T cells* in the biological immune system [17]. The premise of this algorithm is that *normal* data from the training set lead to creating a collection of "self" and generate a set of detectors that sense the "non-self". Anomaly detection is learned by simply "censoring" the detectors until none of them sense the "self", similar to the behaviour of

978-1-6654-0126-5/21/$31.00 ©2021 IEEE

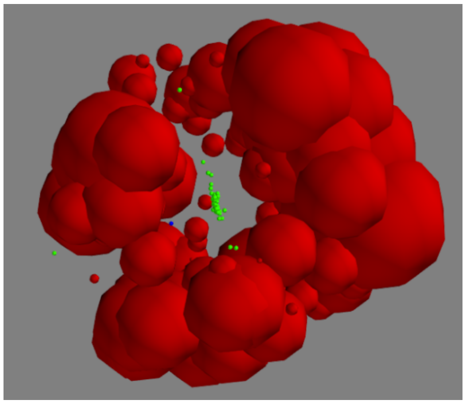

Fig. 2. The NSA algorithm in action, representing the self points with green color, surrounded by detectors with red color. In this case, each self point represents the earbud data of one workout.

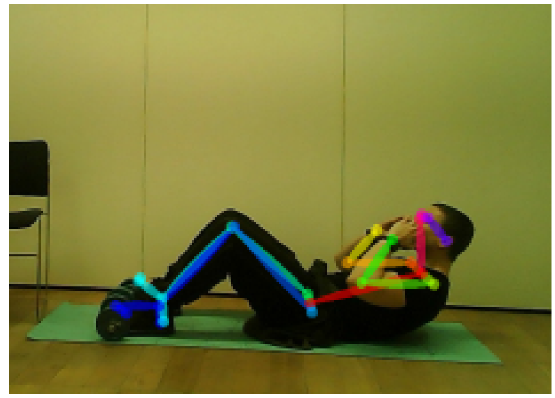

Fig. 3. The complexity of data preprocessing by performing skeleton extraction from raw images.

the immune system. A "monitoring" phase follows where any anomalous data will be detected. This has been used by Forrest et al. [4] to create an intrusion detection system with benefits to computer security. Early work on designing an NSA for real-valued data with "detector spheres" has been performed by Gonzalez et al. [6], and then further improved with spheres of a *variable* radius by Dasgupta et al. [3]. Figure 2 presents a 3D representation of an iteration of NSA. The red spheres are the immune cells that detect anomalies, the green spheres are the accepted signal values and the small blue spheres are anomalies that have managed to escape the view of the detectors. This is a stochastic algorithm, which comes with several good properties when assessing the quality of the estimation.

We adapt NSA to convert the self data into multi-dimensional real-valued points in *Euclidean* space and then generate a set of multi-dimensional *hyperspheres* to detect the non-self.

Our evaluation shows that fault-tolerance can be built into Multimodal Learning with the help of an immune system inspired anomaly detection algorithm (Negative Selection Algorithms). This shows good performance on the MM-Fit dataset, identifying anomalies with high confidence and enhancing robust recognition based on anomaly filtering.

The remainder of this paper is organised as follows. Section II provides a literature review on the concepts used throughout this work, such as multimodal deep learning and Negative Selection Algorithms (NSA). This is followed by an introduction to the project requirements III. The system design is presented in Section IV, with its preliminary evaluation in Section V. The experiment results are discussed in Section VI and conclusions are presented in Section VII.

II. Literature Review

A. Multimodal Learning

Multimodal learning is an expanding area of research, combining several input signals for robust estimations. Strömbäck et al. [18] propose MM-Fit, a solution based on deep neural networks to combine the various sensing modalities of wearable devices and camera for estimating the user physical exercise. This is a form of Human Activity Recognition (HAR). Inertial signals are collected with diverse devices: an earbud (eb_l), two smartwatches (sw_l, sw_r) and a smartphone in the right pocket (sp_r), exploiting their gyroscope and accelerometer sensors. A video camera creates a 3d pose skeleton of the user (fig 3). This data is then processed and combined to create an optimal input for a Convolutional Neural Network. This model accomplishes a maximum accuracy of 96% and comes with its own data set[1]. The data set consists of over 800 minutes of multimodal data from 5 devices built from 20 workouts (of 10 different exercises) across 10 participants.

Similar to [18], Ma et al. [10] uses skeletal data to perform HAR on stroke patients to assess rehabilitation exercises. With this Unimodal approach, [10] achieves 90.9% accuracy on simulated videos. Perhaps this could be increased using MML.

Ordóñez et al. [13] displays a great representation of multimodality. This paper uses MML for the same task as the MM-FIT model, HAR. This paper has yielded a total of a *35%* increase in accuracy over unimodal learning by fusing multiple streams of data input. The Human Activity in question is focused on the actions of assembly line workers in a Skoda car factory. They perform many identifiable actions such as opening car doors. This data set is similar to the MM-FIT data set in many ways yet they have their distinct differences, this paper has used more body sensors than the Smart Gym but lacks a camera/skeleton modal. There does not appear to be any attention to what happens if faults occur within the sensors.

B. Artificial Immune Systems

Forrest et al. [4] are the creators of the Negative Selection Algorithm (NSA). This is inspired by negative selection in biological *T cells*. [4] uses this inspiration to form an antivirus algorithm that uses computer data to represent the self so that a suitable set of detectors can be generated to find anomalies. The results show how powerful NSA's can be, the only

[1] https://mmfit.github.io

drawback found was computational difficulty in generating the detectors. Our paper aims to compose an algorithm that generates a satisfactory set of detectors within a practical time frame.

Dasgupta et al. [3] uses a NASA flight simulator to test aircraft fault detection with a real-valued NSA. The difference between a *real-valued* NSA is that a *hyperdimensional* Euclidean space is used where the self are points and the detectors are *hyperspheres*, detection occurs if a point lies within a detector. The algorithm used involves ensuring detectors don't overlap with each other, moving the detectors in many ways based on what condition they are in and cloning the best whilst also adding random generation of new detectors each iteration. The algorithm terminates when enough of the non-self space is filled. Complicated faults (such as a "Tail 1" failure) are simulated and tested.

Forrest and Dasgupta have worked together to create a tool breakage detection system [2]. The system is successful but required a large data set to generate the self so that the detectors generated would correctly cover enough of the non-self space. Tool breakage detection is not too dissimilar from detecting device faults so we can learn from this paper. The main concern found was that the NSA wouldn't *specifically* detect any particular faults, it would just detect what wasn't *normal*.

C. Other Anomaly Detection Methods

Jayakumar et al. [8] uses *Deep Generalized Canonical Correlation Analysis* to detect anomalies on an earlier version of MM-FIT. This thesis provides robust experiments and shows that anomaly detection for MML is an area to be explored. Multiple *Gaussian mixtures* of various *signal to noise ratios* are added to simulate realistic corruptions. Classification accuracy of the Smart Gym decreased as noise was added. The experiments done in our paper differ from this thesis so it is difficult to correlate the two but evaluating both methods of anomaly detection and seeing if it's possible to combine them could lead to discoveries that move MML forward.

Tadepalli et al. [19] uses machine learning to detect *near-native protein structures*. These structures are anomalous compared to other protein structures so [19] converts the task of supervised learning into anomaly detection. A large data set is used and finding the near-native structures is like finding "a needle in a haystack". Fortunately in our paper, the anomalies are artificially generated so testing the model is not limited by the data set but unfortunately this comes at the cost of a less realistic test environment.

III. Requirements and Analysis

A. Requirements

The goal of this paper is to further the fault-tolerance of MML for HAR. This shall be done by artificially corrupting test input data of one of the sensors and seeing if it is possible to reduce the impact of this corruption. A real-valued NSA will be implemented to detect the corruption and take subsequent action. Evaluation will be done in two ways: testing with and without the MM-FIT model. Tests completed outside of the model will show the detection rate and the best hyper parameters for the NSA whilst tests done within the MM-FIT model will show the improvement made when the NSA is implemented.

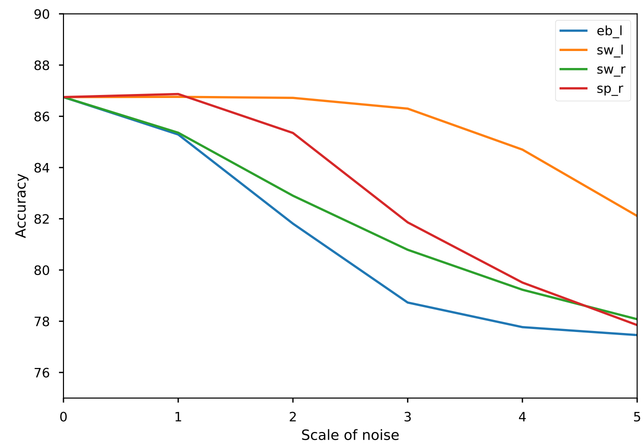

Fig. 4. The impact of noise scale applied to the modalities belonging to one device onto the accuracy of performing the activity recognition. Accuracy decreases as scale (standard deviation) of noise increases on a single workout. Some devices (Smartwatch left/sw_l) are more robust.

B. Analysis

To assess the robustness of MM-FIT, an experiment was run by artificially corrupting the input data of each device and observing the classification accuracy (Fig. 4). There is a negative correlation between accuracy and the scale of noise added, for each device. It is also important to test for false-positives but unfortunately, it is assumed that the test data set contains natural anomalies. Much like how regional accents [20] are not picked up by voice recognition[2], perhaps different people work out in different ways ergo natural outliers in the workout data. These natural anomalies do not affect MM-FIT's classification accuracy (including on unseen test subjects) but to avoid false detections, the NSA requires more steps to accommodate the complexities of the data set used.

To complete this analysis, the requirements of success for this paper are to evaluate an MM-FIT model such that as noise is added to any input device, the classification accuracy does not decrease. Computational efficiency and reduction in complexity is desirable. Another measure of success is to create an NSA with a substantial anomaly detection rate and a minor false positive rate but as discovered in this paper, a seemingly less than adequate NSA can succeed.

IV. Design

The algorithm designed here is a "kissing NSA". This is because the radii of the detectors are made to be close to

[2]Unless the task is to recognise accents [7]

their nearest self point. This gives the algorithm a greedy nature. The idea of "kissing spheres" is loosely inspired by the "kissing number problem" famously debated between Newton and Gregory [11].

A. Design

The design of our Negative Selection Algorithm is a real-valued NSA in one of its primary forms. It is simpler than [3]. It involves a *training* stage[3] to set up the m-sphere detectors and a testing stage which is used to detect anomalies. Principle Component Analysis (PCA) [21] is applied to the training data to reduce to m dimensions.

Training:

$\forall R \in R'$:

$R_c = random(m, B)$

$loop$:

$nS = nearest(R_c, S')$

$R_r = D(R_c, nS_c) - nS_r - k$

$R_c = Tr(R_c, R_r)$ if $\exists S \in S' \rightarrow D(R_c, S_c) < R_r + S_r$

$R = \emptyset$ if $D(R_c, B_c) > B_r$

go to *loop* if $(\exists S \in S' \rightarrow D(R_c, S_c) < R_r + S_r) \wedge (R \neq \emptyset)$

Testing:
For any potential anomaly x, return:

true if $(\exists R \in R' \rightarrow D(R_c, x) < R_r) \vee (D(B_c, x) > B_r)$
false otherwise

Fig. 5. The formulation of the algorithm. This includes the generation of the set of detectors R' and the anomaly detection phase.

Each of these m dimensional points is given a radius to create the self space. A *boundary sphere* is created from the mean centre with a radius double the furthest self point from the centre. A number (nR0) of detectors are randomly generated, their radii are made to *kiss* the nearest self point. [3] uses a similar calculation of detector radius. The reason detectors do not meet the tangent of the self point is because there is concern this may cause the detectors to *overfit*. This could be seen as a *greedy* algorithm because the detectors try to fill as much of the non-self space as possible whilst not being so large that they cause false-positive detections. Any detectors that intersect with the self space are moved in a random direction a distance equal to their radius. Out of bounds detectors are returned to the centre which ends up in them having a radius of zero and rendering them "deleted"[4]. Once all detectors are in bounds and do not intersect with the self space they are considered "censored". Now testing can occur. The potential anomaly goes through the same PCA as the training data. If this transformed piece of data lies within a detector or out of bounds it is considered an anomaly and if it does not it is considered to be part of the self space. *Euclidean* distance is used to measure how far points are from each other.

nR0	number of initial detectors
R'	set of (initially nR0) detector m-spheres
S'	set of self m-spheres derived from data
B	the boundary m-sphere
r	radius
c	centre
m	number of dimensions
k	kissing distance
x	a potential anomaly (any m-dimensional point)
D	distance
Tr	random translation

Fig. 6. The keys for Fig 5. The parameters S_r and k are set to 10 in our evaluation.

B. Finding the Best Hyper Parameters

To find the best hyperparameters, a test was conducted (Fig 9). The number of dimensions is important, too little and there isn't space for detectors to fit in, too many and the data becomes too sparse[5]. An optimal value of 4 dimensions was found. As the number of initial detectors increases so does the detection rate but this does reach an *asymptote* at $nR0 = 100$. Even with 0 detectors, there is a detection rate of $\approx 18\%$. This is due to the boundary sphere. These values were found by artificially adding noise to all workouts to create a test set containing only anomalies. The detection rate is the percentage of anomalies detected by the immune system.

C. Design Extensions

This algorithm is simple so it has many areas open to improvement. The NSA takes an entire workout session as an input but if say, 5 second samples were used, the NSA could act as an alert system for faults in the senses but in such cases safer and better tested algorithms exist [2], [3]. There are many extensions such as comparing the distance of a test point from its nearest self point and detector if it lies in neither [5] or a more rigid selection process could be used [3]. It is possible for detectors to overlap with each other and waste space, especially given the fact that only a limited number of detectors are generated. [3] accommodates this by moving overlapping detectors but this adds run-time complexity. Following the greedy nature of our algorithm, overlapping

[3]As far as NSA's are concerned, "training and testing" are interchangeable with "censoring and monitoring". It is uncertain which phrases are more popular. The author prefers "training and testing" because it can be compared to the equivalent stages in Machine Learning.

[4]This causes the set of detectors to have many empty sets which is not efficient in regards to computational memory.

[5]There exists algorithms tailored for high-dimensional sparse data [9]

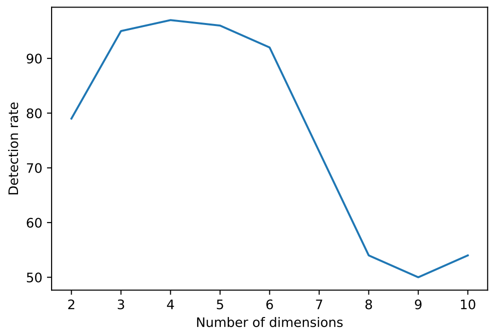

Fig. 7. Number of dimensions

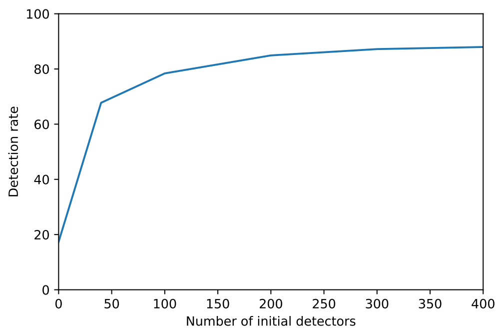

Fig. 8. Number of initial detectors

Fig. 9. How number of dimensions and detectors affects detection rate. These tests were done on separate noisy test sets, both of noise scale 5

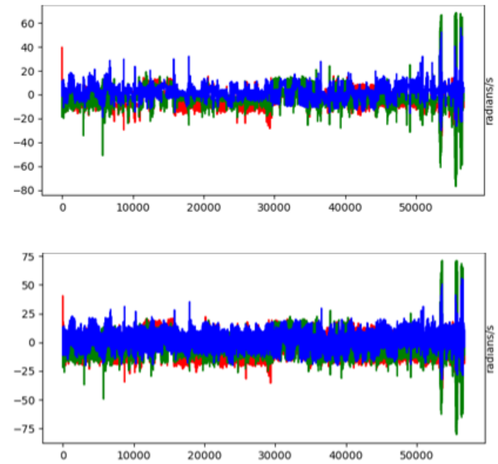

Fig. 10. Adding noise of scale (standard deviation) 3 to a workout modality. Clean above, corrupt below.

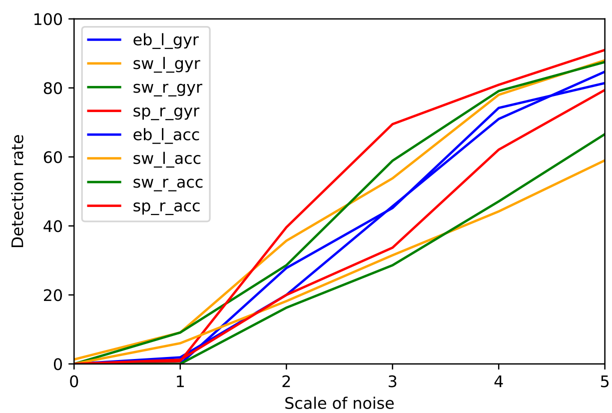

Fig. 11. How scale of noise affects detection rate for each modality. The accelerometer plots are always lower than the gyroscope.

detectors are accepted. Our algorithm only involves spheres, but some methods use ellipsoids [16].

D. Summary

The algorithm will take each workout and transform it into a self point of 4 dimensions via PCA. A *boundary sphere* will be created. 100 *detector spheres* of random location will be generated. These detectors will then be trained to ensure that 1: no self point lies within a detector and 2: no detector lies outside of the boundary sphere. This is done by "deleting" any out of bounds detectors and translating self detecting spheres in a random direction the magnitude of their radius and then updating their radius to *kiss* the nearest self point. Any workout being tested will go through the same PCA as the training data to produce a 4-dimensional point. If this point lies within a detector it is classified as anomalous.

V. Implementation and Testing

A. Implementation

To convert the input workout data into self points, the magnitude of the XYZ values of each sample are treated as feature s and then PCA is applied to reduce dimensionality. For this case, all the workouts were used (including test and validation sets) as NSA training data. This is because the workout data set contains natural anomalies which cause false-positive detections if they are not included. It is assumed that using a larger training data set would decrease these natural anomalies and allow a more robust anomaly detection [2]. An MM-FIT model was trained on all modalities with the standard training and validation set. Two test sets were used, one single workout (to present an example of the NSA in use) and one larger test set with multiple workouts (for reliability). Evaluation on the single test set yielded an accuracy of 87% and the full test set achieved 84% The 3d pose data is very different from the gyroscope and accelerometer data so although it has been included in all tests it has not had noise added to it nor has an immune system been created for it. To see the influence of the 3d pose data, a test was done with only the 3d pose modality which resulted in a test accuracy of 36%.

B. Testing

1) Outside MM-FIT: These tests were used to find the detection rate of the NSA for each modality and to find the best hyperparameters (as mentioned in chapter IV-B). These tests involved producing a self space from all the workout data for each modality and generating an immune system for each. To find the detection rate noise is added to each workout to form a noisy test set and the percentage of this noise set detected became the detection rate. Each test was repeated 10 times and the mean was calculated. Figure 11 provides an example of how different types of data perform differently. The detection rate of the accelerometer component of each device is significantly lower. This test was done by creating a corrupt workout data set using all workouts. It is perplexing to see how low the detection rate is in comparison to the success when implemented to the MM-FIT model. Fault tolerance is observed from scale 2 but this would seemingly be with a detection rate of up to 40%. It is likely that the unseen test subjects set has a higher affinity to anomalies.

2) Within MM-FIT: 5 different scales (standard deviations) of Gaussian normal noise were added to each device in turn. Upon observing the change in data in Fig. 10, it is assumed that corruptions with a noise scale above 3 are rarer. Both gyroscope and accelerometer data of the devices could become corrupted, we have corrupted both at the same time to simulate a total device failure. A unique immune system was generated for each modality and each using 4 dimensions and 100 initial detectors. A different set of immune systems were generated for each experiment. If the immune system detected the noise it would zero out the faulty modality so no noisy input would enter the MML model.

VI. Results and Discussion

The results are a success: MM-FIT is more fault-tolerant when the NSA is implemented.

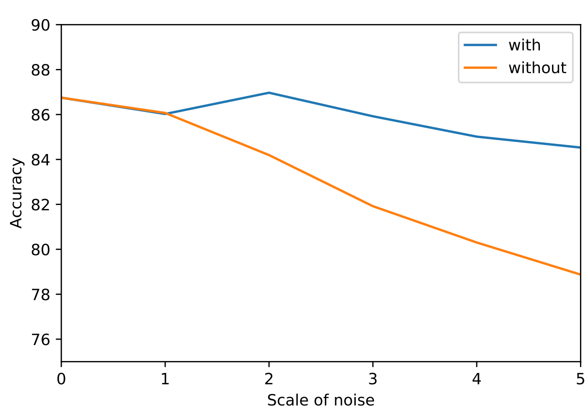

Fig. 12. Mean classification accuracy with and without the immune system on a single workout. The model is more fault-tolerant with the immune system fault detection activated.

When reviewing how the MM-FIT model performs without the immune system (Fig 12), it is hard to ignore how robust the MM-FIT model is to noise (only a 6% decrease in accuracy with the largest scale of noise). As it stands, the MM-FIT model is more fault-tolerant with the immune system implemented.

TABLE I
Classification accuracy over 5 scales of noise.

	0	1	2	3	4	5
eb_l	86.75	85.28	**89.14**	**88.75**	**88.36**	**87.47**
sw_l	86.75	86.68	86.76	86.26	**86.08**	*80.94*
sw_r	86.75	85.34	**86.69**	**86.58**	**86.25**	**85.63**
sp_r	86.75	86.80	85.29	82.10	79.37	**84.08**

In the single workout experiment (Table I) there is less of a negative correlation between accuracy and noise. The values in **bold** show fault-tolerance achieved with the NSA. The impact implementing the NSA had on run-time was found to be *negligible*. The smartphone modality (sp_r) does decline until the 4th scale, this could be because the data is too complicated for the immune systems to detect but once the noise is so great it exits the boundary sphere and becomes detected. The earbud modality (eb_l) increases because the earbud input naturally corrupts the model as a whole so once it is zeroed out the model performs better. In all of these tests, only the gyroscope modalities were detected apart from one case where sw_l_acc was detected on the *5th* scale of noise and the classification accuracy of the MM-FIT model decreased. This is because the model *is* reliant on the accelerometer data so that although it is noisy it is still useful for classifying the correct workout. This also highlights how the data input to the Multimodal model is important. Since removing the earbud input increases accuracy, it might be the case that the optimal input for the workout tested does not include the earbud data. Finding which configuration of inputs regardless if they are corrupt or not could improve the results.

Part of the MM-FIT model training process practices with modalities zeroed out. The intention for this is so that the model can perform well if one of the sensing devices are not available during a workout but coincidentally this is perfect because it means that if a modality has to be zeroed out due to noise the model has already been trained with fewer modalities. Strömbäck et al. [18] discovered that the earbud and smartphone are the less discriminative modalities. This could explain why the model's classification accuracy increases when the earbud data is detected and zeroed out, it simply provides no useful input for this workout. When testing with the larger data set, each device improved when the NSA was implemented.

A. How to Reproduce the Results

To reproduce the results for the larger test set, one must set up the MM-FIT environment with the data set [18]. Then the Python (Pytorch and Scikit) code[6] used for generating the immune systems must be replicated for each modality. An MM-FIT model must be trained with the default parameters but only up to 5 epochs. The example training set, validation

[6]Starter code: https://github.com/Mattias421/nsa-multimodal-machine-learning.git

set and unseen test subjects are selected (resulting in a $\approx 55 : 17 : 28$ split). For each experiment (i.e. device test) a new set of immune systems must be generated. The self space was generated by using all workouts. Each immune system must be 4 dimensional with an initial set of 100 detectors. When testing a device, a Gaussian normal distribution must be added to both the gyroscope and accelerometer components of the device data. The scale/standard deviation of the Gaussian will start at zero and iterate to 5 in increments of 1. As the MM-FIT model is being evaluated the input test data will pass through its prescribed immune system. If the immune system believes it has detected an anomaly the input for that modality will be zeroed out and standard MM-FIT evaluation shall commence and the classification accuracy will be presented. For the single workout test, workout "00" was used.

VII. Conclusions and Future Work

Multimodal Machine Learning (MML) has proven to be an effective solution for Human Activity Recognition. Although deep neural networks are inherently robust to noise, we show how fault-tolerance can be enhanced further with an immune inspired Negative Selection Algorithm (NSA). We implemented this algorithm and specialised it for our task, to detect faults in HAR sensor data. Our prediction pipeline uses the anomaly detection to zero-out the detected faulty inputs and prevent them from corrupting the prediction. Although the NSA may be a greedy algorithm, it is a computationally efficient method of detecting anomalies. The algorithm's low run-time has many benefits, Ngo et al. [12] showing that anomaly detection algorithms can be used on small computing devices.

Future work

A larger test set will give insights into how different users work out and how false-positives could be decreased.

This paper only considers the MM-FIT HAR model, however, other multimodal datasets and tasks exist in the community and we will explore the generalisation of our proposed solution to many of them.

References

[1] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Comput. Surv.*, 41(3), July 2009.

[2] Dipankar Dasgupta and Stephanie Forrest. Tool breakage detection in milling operations using a negative-selection algorithm. Technical report, Citeseer, 1995.

[3] Dipankar Dasgupta, Kalmanje KrishnaKumar, D Wong, and Misty Berry. Negative selection algorithm for aircraft fault detection. In *International Conference on Artificial Immune Systems*, pages 1–13. Springer, 2004.

[4] S. Forrest, A.S. Perelson, L. Allen, and R. Cherukuri. Self-nonself discrimination in a computer. In *Proceedings of 1994 IEEE Computer Society Symposium on Research in Security and Privacy*, pages 202–212, 1994.

[5] Maoguo Gong, Jian Zhang, Jingjing Ma, and Licheng Jiao. An efficient negative selection algorithm with further training for anomaly detection. *Knowledge-Based Systems*, 30:185–191, 2012.

[6] Fabio Gonzalez, Dipankar Dasgupta, and Luis Fernando Niño. A randomized real-valued negative selection algorithm. In *International Conference on Artificial Immune Systems*, pages 261–272. Springer, 2003.

[7] Abualsoud Hanani, Martin J Russell, and Michael J Carey. Human and computer recognition of regional accents and ethnic groups from british english speech. *Computer Speech & Language*, 27(1):59–74, 2013.

[8] Sudharshan Paindi Jayakumar. Robust multimodal deep learning for human activity recognition. In *Master of Science Artificial Intelligence School of Informatics University of Edinburgh*, 2020.

[9] Liping Jing, Michael K Ng, and Joshua Zhexue Huang. An entropy weighting k-means algorithm for subspace clustering of high-dimensional sparse data. *IEEE Transactions on knowledge and data engineering*, 19(8):1026–1041, 2007.

[10] Mengxuan Ma, Benjamin J Meyer, Le Lin, Rachel Proffitt, and Marjorie Skubic. Vicovr-based wireless daily activity recognition and assessment system for stroke rehabilitation. In *2018 IEEE international conference on bioinformatics and biomedicine (BIBM)*, pages 1117–1121. IEEE, 2018.

[11] Oleg Rustumovich Musin. The problem of the twenty-five spheres. *Russian Mathematical Surveys*, 58(4):794, 2003.

[12] DuyHoa Ngo and Bharadwaj Veeravalli. Design of a real-time morphology-based anomaly detection method from ecg streams. In *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 829–836, 2015.

[13] Francisco Javier Ordóñez and Daniel Roggen. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 16(1), 2016.

[14] Keiron O'Shea and Ryan Nash. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.

[15] Valentin Radu, Catherine Tong, Sourav Bhattacharya, Nicholas D Lane, Cecilia Mascolo, Mahesh K Marina, and Fahim Kawsar. Multimodal deep learning for activity and context recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(4):1–27, 2018.

[16] Joseph M Shapiro, Gary B Lamont, and Gilbert L Peterson. An evolutionary algorithm to generate hyper-ellipsoid detectors for negative selection. In *Proceedings of the 7th annual conference on genetic and evolutionary computation*, pages 337–344, 2005.

[17] Timothy K Starr, Stephen C Jameson, and Kristin A Hogquist. Positive and negative selection of t cells. *Annual review of immunology*, 21(1):139–176, 2003.

[18] David Strömbäck, Sangxia Huang, and Valentin Radu. Mm-fit: Multimodal deep learning for automatic exercise logging across sensing devices. 4(4), 2020.

[19] Sivani Tadepalli, Nasrin Akhter, Daniel Barbará, and Amarda Shehu. Identifying near-native protein structures via anomaly detection. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 30–35, 2019.

[20] John Christopher Wells and John C Wells. *Accents of english*, volume 3. Cambridge University Press Cambridge, 1982.

[21] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.