Machine Learning-based Detectors for Magnetic Signal Differentiation applied to Early Cancer Detection

Rita Ramos [1,2,3],
[1]INESC Microsisistemas e Nanotecnologias, 1000-029 Lisbon, Portugal
[2]Instituto Superior Técnico, 1049-001 Lisbon, Portugal
[3] INESC Investigação e Desenvolvimento, 1000-029 Lisbon, Portugal
Corresponding author: rita.ramos@tecnico.ulisboa.pt

***Abstract*—In this work, we disclose an artificial neural network (ANN) classifier for the automatic distinction between labeled targets/analytes and clusters of the labeling magnetic nanoparticles (MNP). The signals are acquired in a magnetic cytometer composed of a disposable cartridge and an acquisition platform. The disposable cartridge employs a microfluidic channel with a 10×100 µm^2 section on top of a silicon substrate where GMR sensors are deposited. We demonstrate that despite the signals from clusters and marked targets intersecting in the feature space if only amplitude and length are considered, an ANN classifier is able to differentiate them if all the samples from the pulse are used as inputs. We explain how the ANN was trained to be generalized for different sampling frequencies. Finally, we explore the relation between classifier accuracy and variables like the number of particles, signal noise, and maximum channel height.**

***Index Terms*—Magnetic Cytometry, Low noise, Machine Learning, Oversampling**

I. Introduction

A. Magnetic Flow Cytometer

The magnetic flow cytometer (MFC) is used to count a specific analyte in which a sample is present. For example, to count and identify cancer cells in a blood sample, these analytes are the cancer cells. But since a blood sample does not contain magnetic nature, the cancer cells cannot be identified. Therefore, a process, through which the sample must pass, is necessary, so that analytes became magnetic.

In this process, magnetic nanoparticles (MNP) are mixed in the solution and will bond to the specific marker. Depending on some factors, like efficiency, the number of target cells contained in the sample, and others, some or several MNP may remain free in the sample. The resulting mixture contains the original solution with the marked cells and free MNP. This free MNP that did not bind to targets of interest, end up joining forming clusters.

B. Cluster's Problem

After the sampling process described before, the sample can be introduced in the MFC, and the magnetic field produced by the magnetic particles is gradually picked up by the magnetoresistive sensors (MR) sensors (Figure 1).

Since, in addition to the labeled analyte being present in the sample, clusters are also present, and thus the resulting magnetic field also contains data from the clusters. This interference of free magnetic labels on the detection signals of magnetically labelled targets may lead to a false positive. For this reason, it is important to try to differentiate both signals as much as possible.

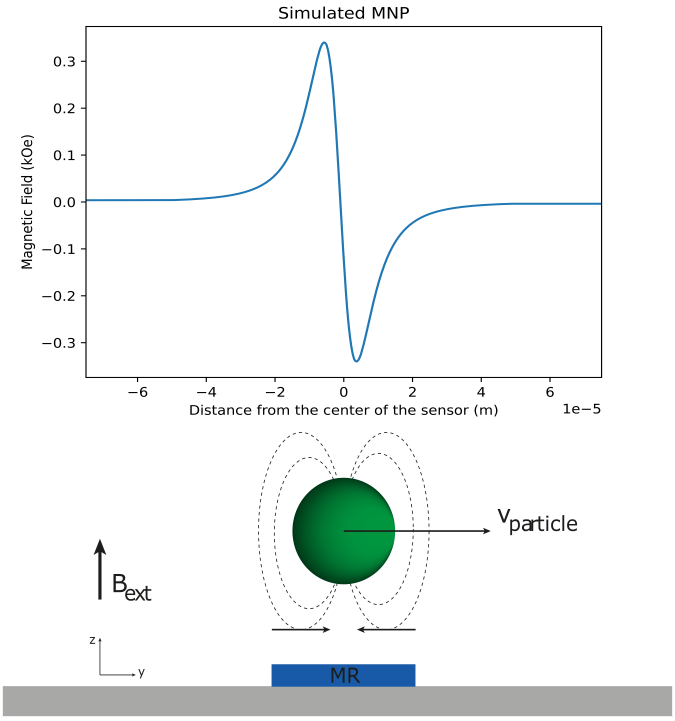

Fig. 1. Interaction between the MR sensor and a particle that crosses the channel of a cytometer.

II. Simulations

A. MNP Simulation

It is known that the signal of the magnetic field of the magnetic particles detected by the MR sensors in the cytometer has a specific signature and that it resembles the Gaussian pulse [Figure 2], or its derivates depending on the particle magnetization angle.

With the advancement of some works [1] [2], it was possible to assume that an MNP can be modeled using the simulation of

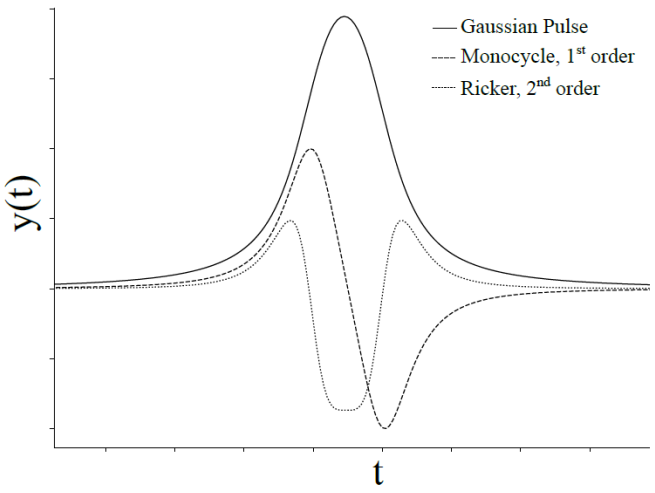

Fig. 2. Illustration of Gaussian Pulse and its derivates.

the magnetic dipole. The equation 1 represents the integral of the magnetic field [in Oe] produced by the magnetic dipole in the sensitive axis (x) of the sensor used in this simulation. x_0 and y_0 represent the center of the sensor, l and w represent the length and width of the sensor, θ represents the magnetization angle, and x and y represent the position of the center of the dipole relative to the center of the sensor.

$$Hx = \frac{\left(\int_{-\frac{1}{2}-y_0}^{\frac{1}{2}-y_0} \int_{-\frac{w}{2}x_0}^{\frac{w}{2}x_0} \frac{M}{4\pi} \frac{4\pi}{1000} \times \frac{3\cos\theta(x+h)+\sin\theta x^2}{(x^2+y^2+h^2)^{5/2}} - \frac{\sin\theta}{(x^2+y^2+h^2)^{3/2}}\, dx\, dy\right)}{w \times l} \quad (1)$$

The simulation of an MNP with 2 μm of diameter is represented in Figure 3a). It is possible to see that the simulation of a particle produces a signature close to what is a Gaussian monocycle pulse, as mentioned before. Figure 3b) represents the simulations for a particle traveling in the center of the channel with the height's parametric sweep range from 3 to 15μm, relative to the plane perpendicular to the channel's base, and direction of flow.

B. Cell Simulation

Based on the confirmation that the dipole simulation can be used to describe the comportment of an MNP, it is now possible to simulate the MNPs around a cell.

Through Figure 4, it is possible to observe the result of the simulation that uses a sphere to represent a cell with 10 μm of diameter and an algorithm was performed that randomly distributes points along the spherical surface (blue dots), thus representing the MNPs distributed around the cell.

Using this cell position simulation combined with the MNP simulation described in section II-A, it is possible to simulate a cell, as it possible do see from Figure 5 that shows the simulation of 15 particles randomly arranged on the spherical surface of a cell, as well as the signal resulting from the sum of all particles. The cell has 10 µm of diameter and is located at a height (z) of 6 µm.

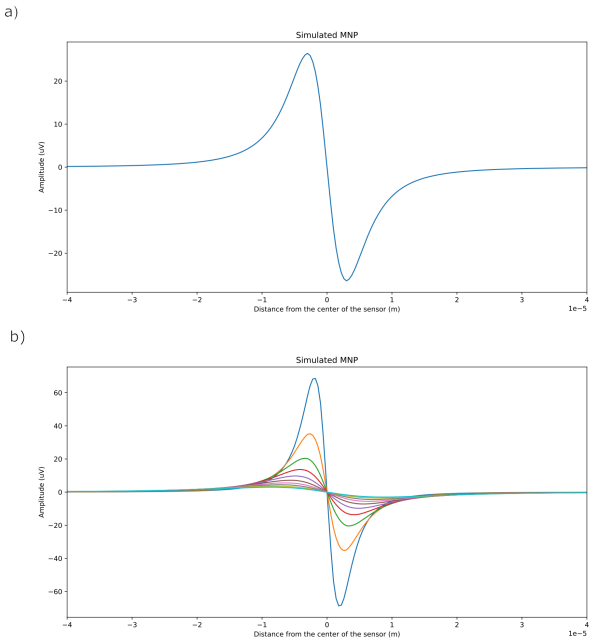

Fig. 3. Representation of the simulation of a single MNP: a) 1 MNP at 5 µm of height; b) 1 MNP at 10 different heights from 3 µm to 15 µm.

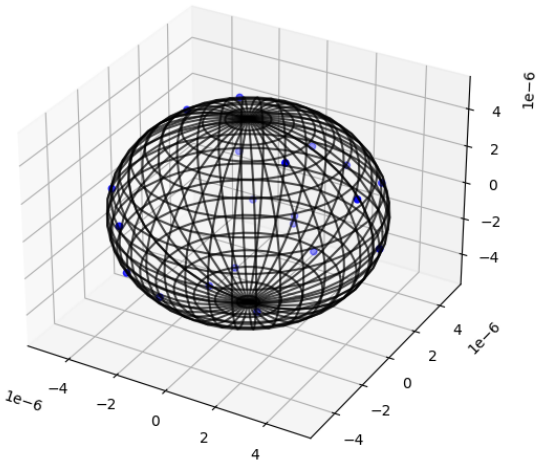

Fig. 4. Simulation of a cell with 10 μm of diameter and 20 particles (blue dots) on its spherical surface randomly distributed.

It is possible to observe that the MNPs are at different distances from the center of the sensor, to the x-axis, and as they are also located at different heights (z-axis), the MNPs that are closer to the sensor have a bigger magnetic field than those that are further away.

C. Cluster Simulation

Clusters are sets of MNPs that never came together with the analytes/cells present in the samples, and that, therefore, end up joining each other.

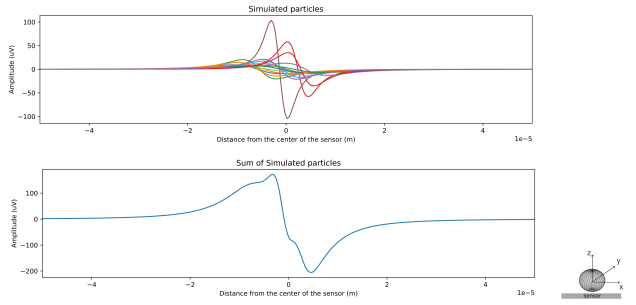

Fig. 5. Simulation of 15 MNP around a cell with 10 μm diameter, located at 6 μm of height. Above, the representation of each particle, below, the representation of the sum of the simulated particles.

To simulate the clusters, it is necessary to define how they are grouped. Assuming that MNPs are spheres, it is possible to use the packing of spheres as a grouping method. In geometry, a sphere packing is an arrangement of non-overlapping spheres within a containing space. The spheres considered are usually all of identical sizes, and the space is usually three-dimensional Euclidean space. There are two main methods of sphere packing, square packing, and hexagonal packing. Both are represented in Figure 6.

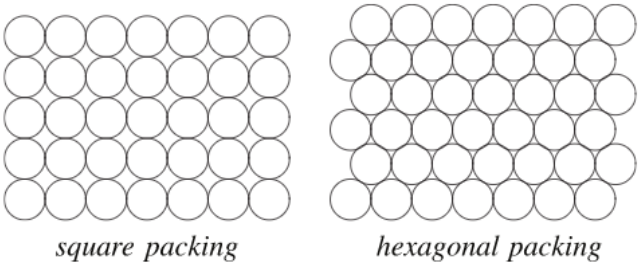

Fig. 6. Square Packing representation for circles (MNPs).

The chosen method used to simulate clusters was square packing, because, although it is a method that uses more free space than hexagonal packing, it is easier to implement, since if we define a main sphere/MNP, with 2 µm of diameter, placed in a certain position, all the others are situated at a distance of 2 µm in all x, y, z directions from the main one.

In Figure 7 is represented the simulation of a cluster with 15 MNP using the square packing.

It is possible to conclude that, unlike cell simulation, in cluster simulation, the MNPs, from the point of view of the x-axis, are more close to each other and they have all spaced the same distance, which in this case is 2 µm which corresponds to the diameter size of the particle. This happens because in this case there are no cells between the particles, and therefore they are closer to each other. For this reason, it is natural that the signals from the cluster simulation have a greater amplitude because as the distance between the particles is smaller, the sensor detects a higher magnetic field when the particles pass through it.

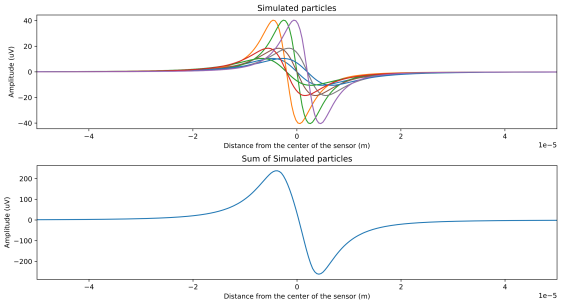

Fig. 7. Cluster Simulation located at 6 μm of height for 15 particles using square packing.

III. Simulation Results

A. Cell vs Cluster Simulation

One way to compare and differentiate the signals from the simulation of cells and clusters is to observe the two signals simultaneously when subjected to the same conditions of particle numbers and heights in relation to the sensor's plane. The following Figure 8 shows the simulated signals of cells and clusters for the same number of particles (15 MNP) with a height (from the center of the cell/cluster to the plane of the sensor) relatively close to the sensor (6 µm).

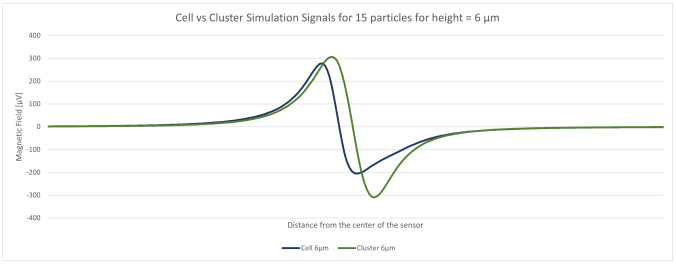

Fig. 8. Cell vs Cluster Simulation signals for 15 MNP located at 6 μm of height to the plane of the sensor.

From this Figure 8 it is possible to conclude that the magnetic field signal coming from the cluster simulation (green signal) has a greater amplitude in comparison to the cell simulation (blue signal). To compare both signals with a different perspective, Figure 9 represents the same signals but for a distance further from the sensor plane (10 µm).

It is possible to observe that the smaller the distance in height of the particles to the sensor's plane, the greater the amplitude of the signal. The reason is because the magnetic field produced by the particles, felt by the sensor, is stronger, once the magnetic particle is closer to the center of the sensor. It is possible to conclude that the simulations seem to correctly demonstrate what happens in reality.

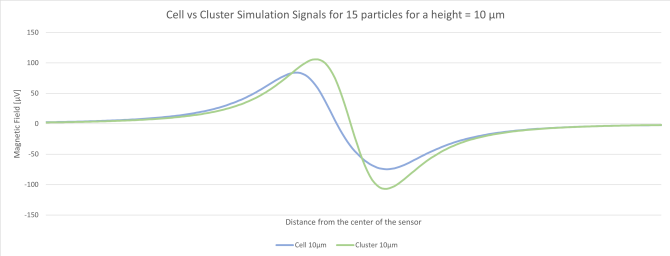

Fig. 9. Cell vs Cluster Simulation signals for 15 MNP located at 10 µm of height to the plane of the sensor.

IV. Adding Noise to Simulations

Although the simulation results seem acceptable and similar to what is expected from the real signals, the absence of the noise signal in the simulation can be a negative point for the artificial neural network's results, because the signal resulting from the interaction of the magnetic particles with the MR sensors in the cytometer contains noise. The simulated results shown here have no noise, which makes it very simple for the neural network to distinguish the signals between cell and cluster, so the ANN's results would be inconclusive, as they would not translate to real data.

Thus, to add noise to the simulation, a value obtained randomly through a normal distribution was added to the original signal. A normal distribution (also known as Gaussian, Gauss, or Laplace–Gauss distribution) is a type of continuous probability distribution for a real-valued random variable. The general form of its probability density function is described in Equation 2 where μ is the mean or expectation of the distribution and σ is it standard deviation. For this case, it can be assumed that μ is zero and σ will be the parameter that defines the amplitude of the noise signal in µV.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}, \tag{2}$$

Figures 10 and 11 represent the simulation for 1 MNP located at 5 µm of height in relation to the center of the sensor with a noise signal of σ=2 µV and σ=10 µV, respectively.

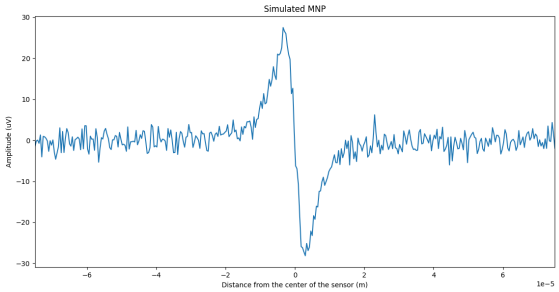

Fig. 10. Simulation of an MNP located at 5 µm of height with a noise signal of σ = 2 µV.

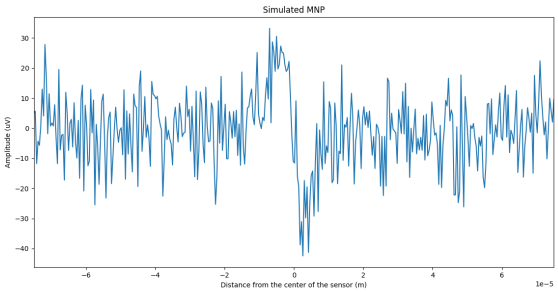

Fig. 11. Simulation of an MNP located at 5 µm of height with a noise signal of σ = 10 µV.

By observing the figures, the greater the value of the sigma parameter, the greater the amplitude of the noise signal, and therefore, the more difficult it is to distinguish the signature from the presence of the MNP in the microfluidic channel of the cytometer. The same will also happen in relation to the height that the particle is to the sensor, because for the same sigma value, the greater the distance to the sensor's center, the more difficult it will be to find the signature. This can be observed by comparing Figures 10 and 12, the difference between them being the height of the MNP. For particles located further away from the sensor, the amplitude of the signal is smaller, as was concluded before, and it becomes more difficult to detect their presence.

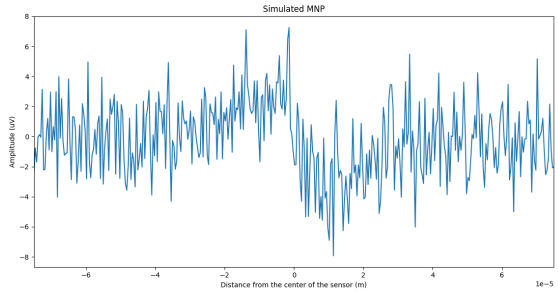

Fig. 12. Simulation of an MNP located at 15 µm of height with a noise signal of σ = 2 µV.

Thus, concluding that the implementation of cell and cluster simulations is well defined and according to what happens in reality, it is now possible to use the data obtained from the simulated magnetic field to train the artificial neural network.

V. Artificial Neural Networks

A. Training DataSet

A training data-set is the base for any machine learning application, and it contains the data that will be used by a ML model, in this case, an ANN, to train and predict the results. Thus, remembering that the objective is to distinguish between a cell and a cluster through the signal of the magnetic

TABLE I
REPRESENTATION OF THE DATA-SET TRAINING.

			ANN Output	ANN Input
NºParticles	Sigma's Noise Signal [μV]	Height [m]	Cell/Cluster	Magnetic Field [μV]
17	2	$5e^{-6}$	1	[-1.94703, 1.90645, ... , 0.86984, 1.33735]
17	2	$5e^{-6}$	0	[1.05238, 3.31398, ... , 1.82329, 3.76494]
17	3	$10e^{-6}$	1	[-0.59911, 2.35205, ... , 1.44749, -0.91202]
17	3	$10e^{-6}$	0	[-2.63580, -1.91170, ... , 0.74629, 1.26064]
(...)				
70	12	$15e^{-6}$	1	1.11419, 3.08200, ... , 0.41205, -0.05114]
70	12	$15e^{-6}$	1	[-1.15678, -1.056711, ... , 0.75632, 2.27631]

field, we have as input of the neural network the signal itself, that is, the signature that was shown earlier that resembles the Gaussian pulse. So this task can be approached as a classification problem, and as an output of the ANN, we have the value '1' if this signal corresponds to a cell and the value '0' if it is a cluster. Each input has 250 points/samples that correspond to the pulse and that is translated by the sum of the magnetic fields produced by the magnetic particles that constitute the cell or the cluster. Table I is a representation of how is the data-set that will serve as training for the neural network.

Table I only represents a few cases for information purposes, however, the dataset is made up of thousands of cases since each characteristic that affects the magnetic field will vary as follows:

- Height ranges from 5 to 15 μm with a step of 0.1 μm;
- The Sigma value of the noise signal varies between 2-12 μV with a step of 1 μV;
- The number of particles per cell/cluster will take the following values: [6, 10, 17, 22, 31, 40, 55, 62, 70].

The choice of the number of particles was taken into account that it would be very difficult to make a dataset with all possible numbers of particles, and for that reason, a random number was chosen for every ten up to a maximum of 70 particles. Furthermore, for each different case, 5 different noise vectors are generated, to take into account that each noise vector is implemented randomly, through the equation 2, and to have more data in the dataset.

B. ANN Structure and Training

An ANN consists of 3 main layers: the input layer, the hidden layer, and the output layer. Using the same number of neurons in the input layer for the hidden layer, which in this case is 250, which corresponds to 250 points/samples of the magnetic signal obtained through the simulation, the accuracy results obtained were the best compared to other values or methods. Since the input is fixed, because it always has 250 values extracted from the signal, we can assume that the number of neurons in the input and the hidden layer is always fixed and equal, as we have obtained the best accuracy results in this case. The output layer is composed of only one neuron once this is a classification problem where there are only two classes, 'cell - 1' or 'cluster - 0'. For that reason, the activation function in this layer is the sigmoid function. The activation function, for the hidden layer, is the ReLu function because in addition to its simplicity, this function has a strong biological motivation, and it has been demonstrated to enable better training of deeper networks than other activate functions. The data preparation, the ANN architecture, and training were performed in python language, using the Keras library with the TensorFlow 2.2.0 back-end and the sci-kit learn package.

Training the data-set with 150 epochs, the maximum value of accuracy was 99.68%, which means that, with a data-set of 77 thousand cases, 247 cases cannot be correctly classified by the model. Accuracy means the degree to which the result (of the model to predict if the signal is a cell or a cluster) is correct. The objective of this project is to distinguish between cell and cluster magnetic signals, for a different number of particles and positions. So to interpret better our accuracy results, the model was trained using the entire dataset, and to observe the predictions of the model as a function of different parameters, the data-set test was split into a different number of particles and different heights to classify and observe how the accuracy of the model varies with the variation of these parameters. The results obtained are represented in the next section.

VI. RESULTS

The first results obtained are presented in Figures 13 and 14. Figure 13 represents the accuracy of the model as a function of 2 parameters: the number of particles per cell/cluster, and sigma value of noise signal. For a bigger number of particles, the color is darker and for smaller particle numbers the signal is lighter. Figure 14 represents the accuracy of the model as a function of the height of the cell/cluster to the plane's sensor and also the sigma value. For heights closer to the sensor (5 μm) the signal is darker and for heights more furthest (15 μm) the signal is lighter.

For larger particle numbers and smaller heights the magnetic field is higher since in the first case the magnetic field produced is larger because there are more particles, and in the second case because as we are closer to the sensor, the signal is

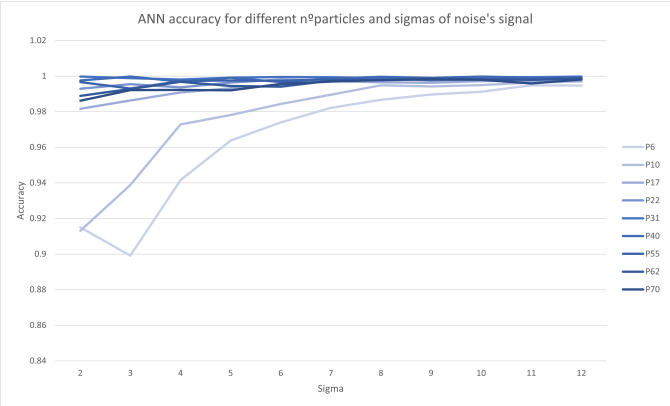

Fig. 13. ANN accuracy for different nºparticles and sigma values.

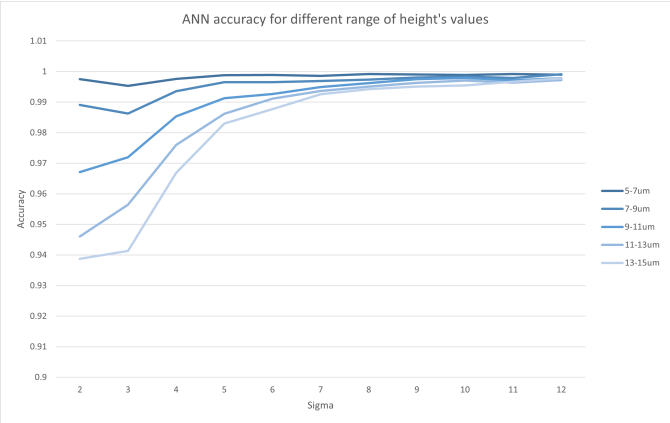

Fig. 14. ANN accuracy for different range of height's values and sigma.

felt by it more intensely. Furthermore, for higher sigma values, the noise signal is also higher. Putting these two facts together, supposedly what should happen was that the accuracy should decrease over sigma values since the noise signal was higher and it should be more difficult to distinguish between cell and cluster in the middle of the noise. However, what happens is precisely the opposite, which leads us to the conclusion that the ANN, from a certain moment on, no longer distinguishes between cell and cluster, to start distinguishing between noise signal and cluster signal (for having the largest amplitude of magnetic field), which is a bit easier to differentiate. This is a negative result because the ANN begins to misclassify with a lower SNR values (greater sigma values), which will be more common in reality.

That said, for this not to happen, it was decided to decrease the maximum number of sigma of the noise signal to 7 µV, since it was from this sigma value that accuracy became very high and stable close to 1, as can be seen from the previous figures. From now on, to make the results more understandable, these will be represented in SNR values in dB instead of sigma values, following Equation 3. Besides that, the number of particles less than 10 can be removed because the signal is not strong enough to make useful conclusions.

$$SNR\,[dB] = 20log\left(\frac{\sqrt{\left(\sum_{n=1}^{n=250} Magnetic\,Field(n)\right)^2}}{\sqrt{\left(\sum_{n=1}^{n=250} Noise\,Signal(n)\right)^2}}\right) \quad (3)$$

Applying the decisions made above, the result that translates the accuracy as a function of SNR value intervals by the number of particles is represented in Figure 15. In this case, the noise signal is higher for lower SNR values, so the results show that, on average, with increasing SNR value range, accuracy tends to improve. Furthermore, for larger numbers of particles that make up cells and clusters we have higher accuracy, to smaller numbers of particles. Figure 16 is represented the accuracy of the model as function of the height of the cell/cluster to the plane's sensor and also SNR value intervals. For smaller height's values, when the position is more closer to sensor, the accuracy is greater. For bigger distances to the sensor the accuracy decreases. The Figure shows that overall there is a tendency for accuracy to improve as SNR values increase.

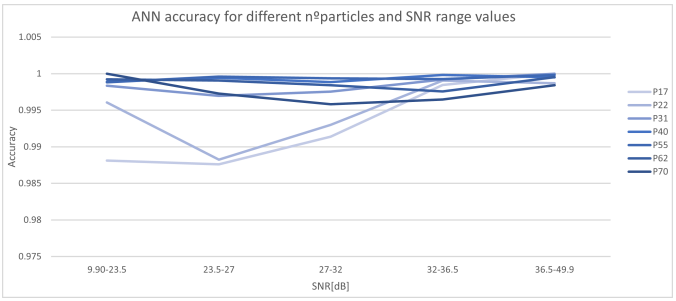

Fig. 15. ANN accuracy for different nºparticles and SNR range values.

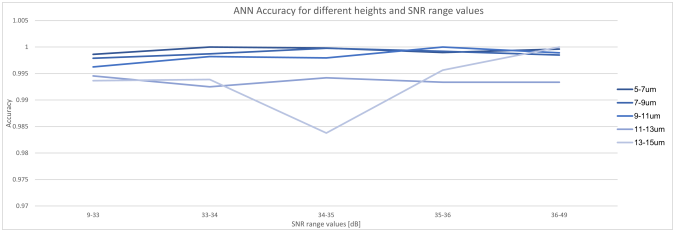

Fig. 16. ANN accuracy for different heights and SNR range values.

It can be concluded that this ANN can differentiate, with good accuracy, the cell signals from cluster signals for a system with a maximum RMS noise signal of 7 µV which is the equivalent of a minimum SNR value of 9.90 dB, for cells and clusters with a minimum of 20 particles. Although there is no maximum number for particles, the maximum number used to train was 70 because it seems to be a reasonable number for particles that are attached to a cell. The actual value of RMS noise in the MFC that we are working on in this project is between 2-5 µV at 200KHz bandwidth, so this model can predict any signal that is between these values.

To have a visual perception of the signals that we are talking about, the following Figures 17, 18, 20 and 19 represent both signals (cell/cluster) in the same conditions to exemplify the magnetic signals with the maximum and minimum values of SNR that this model can predict. By observation of the first to Figures (17 and 18), it can be concluded that for a minimum SNR value, that corresponds to worst conditions, both simulations look similar which can be translated as harder for the ANN to distinguish the signals. Another characteristic that must be mentioned is that for the same conditions, the cluster simulation has always a bigger SNR value, that happens because the cluster signal has a bigger amplitude (because of all the reasons mentioned before) and the noise signal has the same amplitude for both signals so SNR values will be bigger for cluster simulations. For better conditions (Fig. 20 and 19), the cluster simulation has a greater amplitude than cell simulation so it is much easier for ANN to differentiate it.

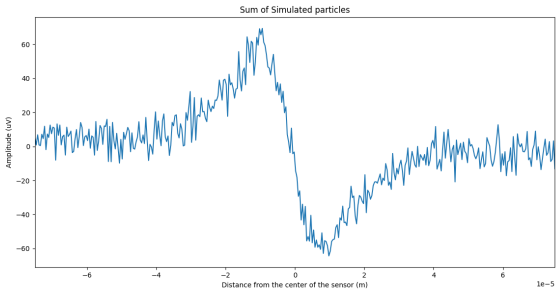

Fig. 17. Example of a signal for minimum SNR (11.41 dB) for a cell simulation with 20 particles at 15 μm of height.

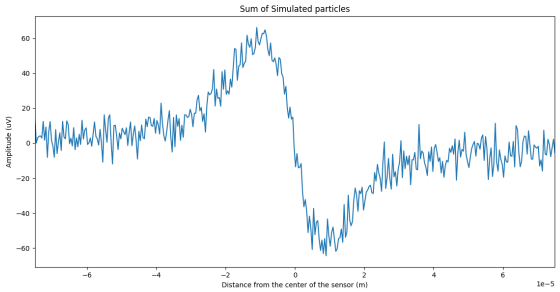

Fig. 18. Example of a signal for minimum SNR value of 13.24 dB, for a cluster simulation with 20 particles at 15 μm of height.

VII. Signal Subsampling

The results obtained so far translate the accuracy of the model that uses as input the signal that results from the sum of the magnetic field produced by the particles. This signal contains 250 samples, however, it will not always be possible to obtain this amount of samples, because if the particles travel

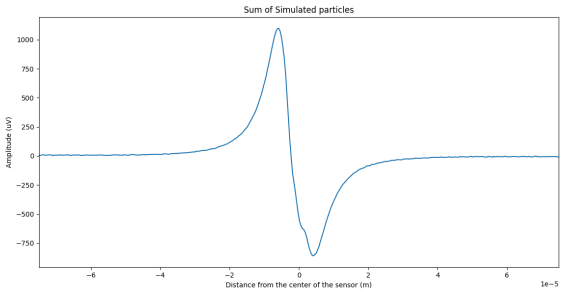

Fig. 19. Example of a signal with a SNR value of 44.9 dB, for a cell simulation with 70 particles at 5 μm of height.

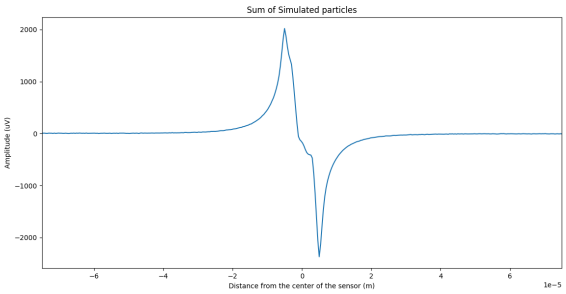

Fig. 20. Example of a signal with a SNR value of 48.02 dB, for a cluster simulation with 70 particles at 5 μm of height.

at a higher speed for the same sampling frequency we will have a signal with fewer samples. So the idea is to try to understand if the artificial neural network, that was implemented, can distinguish the signals when the particles travel at a higher speed. And if so, establish a limit situation by which the model fails to distinguish.

In Figures 21, 22, and 23 there are some results of ANN accuracy for simulations with different particle's speed and with a frequency sample of 200 kHz.

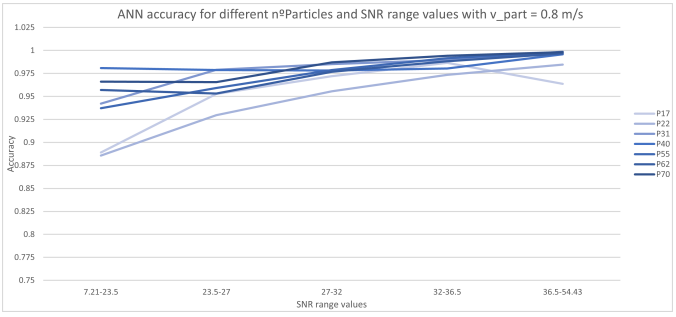

Fig. 21. ANN accuracy for different nºparticles and SNR range values for a particle speed of 0.8 m/s, which corresponds to a signal with 30 samples.

For observation of the figures, it is possible to conclude that for a greater particle's speed (fewer samples), in general, the

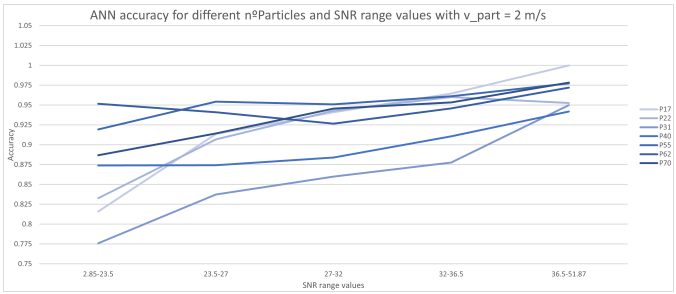

Fig. 22. ANN accuracy for different nºparticles and SNR range values for a particle speed of 2 m/s, which corresponds to a signal with 14 samples.

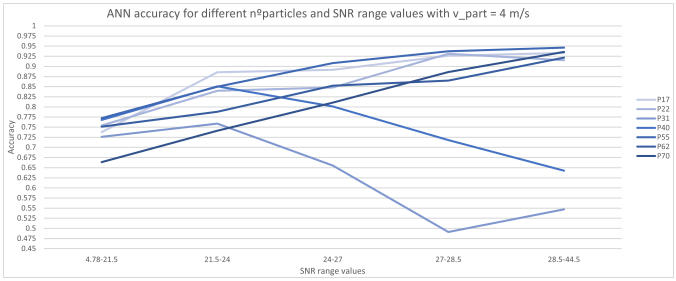

Fig. 23. ANN accuracy for different nºparticles and SNR range values for a particle speed of 4 m/s, which corresponds to a signal with 6 samples.

accuracy values tend to decrease, as it is possible to see from Figure 24, which shows how the model accuracy varies as a function of the particle's velocity.

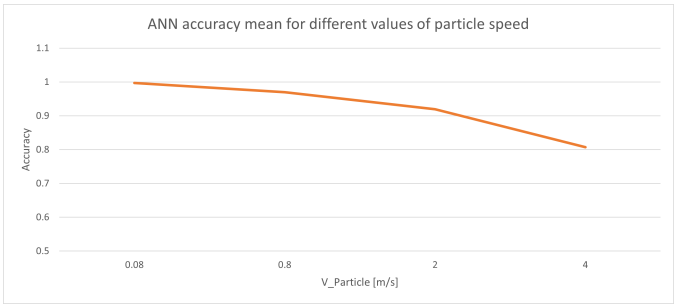

Fig. 24. ANN accuracy mean for different values of particles speed.

According to [3] there is a estimation that is necessary at least 10 samples to fully characterise the pulse, as suggested by [4], [5]. This estimate arises through the Equation 4, that estimates the maximum flow rate in µl/min to use on the cytometry experiment considering: sampling rate (F_s) the ratio between pulse duration and maximum to minimum duration ($\alpha = T_{full}/T_{H_L}$) empirically estimated to be between 3-4 in the case of bipolar Gaussian mono-cycles, the sensor length (L_{sens}), the minimum desirable number of samples to reconstruct the signal (S_{min}), the microfluidic channel's height (h_c) and width (w_c), as well as the channel section (A_{sec}).

$$Flow\ rate = \frac{L_{sens} \times \alpha \times F_s}{\left(-0.56\left(\frac{h_c}{w_c}\right)^2 + 1.15\left(\frac{h_c}{w_c}\right) + 1.5\right) S_{min}} A_{sec} \times 60 \times 10^9 \quad (4)$$

Thus, according to this and by observing the results present in Figure 24, it is possible to conclude that using at least 10 samples (that corresponds to approximately 2.5 m/s of particle's speed) the ANN can differentiate the signals with an overall good accuracy.

VIII. Conclusions

By observing the accuracy results obtained, it was concluded that this model can differentiate the signals with greater precision for a system with a maximum RMS noise signal of 7 µV, for cells and clusters ideally with a minimum of 20 particles. By subsampling the signal it can also be concluded that the maximum speed that the particles can travel in the microfluidic channel is at a speed of 2.5 m/s.

References

[1] J. Loureiro, P. Z. Andrade, S. Cardoso, C. L. da Silva, J. M. Cabral, and P. P. Freiras, "Magnetoresistive chip cytometer," Labo on a chip, vol.11, no. 2255, 2011.

[2] J. Loureiro, R. Ferreira, S. Cardoso, P. P. Freitas, J. Germano, C. Fermon, G. Arrias, M. Pannetier-Lecoeur, F. Rivadulla, and J. Rivas "Toward a magnetoresistive chip cytometer: Integrated detection of magnetic beads flowing at cm/s velocities in microfluidic channels," Applied Physics Letters, Vol. 95, no. 3, p. 034 104, 2009.

[3] D. M. Caetano, "Circuits and signal processing for magnetoresistive sensor arrays", Ph.D. dissertation, 2021.

[4] D. Issadore, J. Chung, H. Shao, M. Liong, A. A. Ghazani, C. M. Castro, R. Weissleder and H. Lee, "Ultrasensitive clinical enumeration of rare cells ex vivo using a micro-hall detector", Science Translational Medicine, Vol. 4, no. 141, 2012.

[5] D. Issadore, H. J. Chung, J. Chung, G. Budin, R. Weissleder and H. Lee, "Micro-hall chip for sensitive detection of bacteria", Advanced Healthcare Materials, Vol. 2, no. 9, p. 1224-1228, 2013.