Detection and Classification of Supernova Gravitational Waves Signals: A Deep Learning Approach

Authors[1][2]

[1] Department of Applied Physics, Fukuoka University, Nanakuma 8-19-1, Fukuoka 814-0180, Japan
[2] SUPA, School of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, UK

We demonstrate the application of a convolutional neural network to the gravitational wave signals from core collapse supernovae. Using simulated time series of gravitational wave detectors, we show that a convolutional neural network can be used to detect and classify, based on the explosion mechanisms, the gravitational wave signals buried in noise. Our results suggest that for magnetorotational core collapse supernovae events at 50 kpc with a network of advanced LIGO, advanced VIRGO and KAGRA, our convolutional neural network can achieve a true alarm probability of 85% at false alarm probability of 0.1. For neutrino-driven supernova events, a true alarm probability of 76% is achieved for sources at 10 kpc. The true alarm probabilities increase to 95% for signals of magnetorotational mechanism at 50 kpc and 92% for those of neutrino-driven mechanism at 10 kpc if the detector network is consisted of LIGO A+, advanced VIRGO and KAGRA, all at false alarm probability equal to 0.1.

I. INTRODUCTION

Since 2015 when LIGO has made the first direct observation of gravitational waves (GWs) from the merger of a binary black hole[1], there have been numerous observations of GWs from similar systems in the observation runs of LIGO and VIRGO [2–4]. These discoveries have been a crucial milestone in GW astronomy and have opened up a new window on the sky. More recently, LIGO and VIRGO have observed the GWs from a binary neutron star merger[5–7]. For this event, the GWs and the associated gamma-ray burst were observed simultaneously. Other counterparts across the electromagnetic spectrum were also observed by later follow-up observations [7]. In the near future, many more observations of GWs from similar compact binary coalescence systems can be expected as KAGRA starts joint observations with LIGO and VIRGO [8–10].

In addition to compact binary coalescences, massive stars with $10 - 100 M_\odot$ at zeros-age main sequence ending their lives by becoming core collapse supernovae (CCSNe) are also considered to be potential sources to the second generation detectors such as the advanced LIGO (aLIGO) [11], advanced VIRGO (AdVirgo) [12] and KAGRA interferometers [8, 13, 14]. It is currently not entirely clear to astronomers how such a massive star becomes supernova. The basic theory of the explosion, confirmed by the neutrino events observed from SN1987A[15], goes as the following. A massive star at the final stage of its life will form core that is composed of iron nuclei after it has burned all its stellar fuel via nuclear reaction. The iron core is supported by the pressure of relativistic degenerate electrons. If the mass of the core exceeds the effective Chandrasekhar mass[16, 17], core collapse will ensue and continue until it reaches nuclear density. The nuclear equation of state will then stiffens by the strong nuclear force above the nuclear density and stops the core collapse. The inner core will bounce back

and a shock wave will be sent through the infalling matter. By losing energy to the dissociation of the iron nuclei and to neutrino cooling, the shock wave will stall. For the star to become a supernova, the shock wave will need to be revived[18]. However, the mechanism via which the shock wave is revived and causes the explosion has been an unsolved problem and the subject of intense study.

There exist two most popular theories, the neutrino-driven mechanism[17, 19] and the magnetorotational mechanism[20–22]. For supernova progenitors with core rotation too slow to affect the dynamics[23, 24], the neutrino-driven mechanism is believed to be the active mechanism. The majority of the observed CCSNe can be explained by the neutrino mechanism[25]. The neutrino mechanism[19, 26] suggests that about $5 - 10\%$ of the outgoing neutrino luminosity is stored below the shock, which causes turbulence to occur and thermal pressure to increase. The stalled shock can be revived by their combined effects[27]. Producing a CCSN via the neutrino mechanism may also require convection and the standing accretion shock instability[28]. On the other hand, the magnetorotational mechanism requires rapid core spin and strong magnetic field[29–33]. Together, they may produce an outflow that may cause the most energetic CCSNe observed. The magnetorotational mechanism may be able to explain the extreme hypernovae and the observed long gamma-ray bursts[34–36].

Correctly classifying the GW from a CCSN is important in understanding the explosion mechanism. As GWs are emitted in the central core of a CCSN, they are likely to carry direct information of the CCSN and therefore provide a probe of the explosion mechanism that produces them. In GW astronomy, when it comes to the search for signals from compact binary coalescences, the established routine is matched filtering. However, as the emission process of the GWs from CCSNe is affected by turbulence in the post-bounce and is stochastic in nature, the signal evolution cannot be

predicted robustly[37–39]. This in turn prevents the match filtering routine to be applied to CCSNe. Methods and algorithms have been developed for the detection and classification of signals from CCSNe. For example, a method known as principle component analysis has been developed[40–44]. This method creates a set of component basis vectors from a set of CCSN waveforms of a particular mechanism that represent the common features of the waveforms of that mechanism. There have been other approaches developed in the literature such as Bayesian inference[41], Bayesian model selection[45] , multivariate regression model[46], maximum entropy[47], maximum likelihood[48] and Tikhonov regularization scheme[49, 50].

In recent years, the field of machine learning and its sub-field, deep learning, have been rapidly developing because of its potentials in many fields [51–56]. For example, deep learning have been successfully applied to fields including medical diagnosis[57], object detection[58], image recognition/processing/generation[51, 59–61], and language processing[62]. In GW astronomy, deep learning has mostly been applied to both the identifications of glitch[63–66] and signals[67, 68]. Convolutional neural network (CNN) is a deep learning algorithm that has the advantage of capturing spatial and temporal features of the input data. Another advantage of the use of a CNN in the detection of a signal is that a CNN is relatively computationally cheap compared to other more traditional methods. This is because the heavy computational work is usually done during the training stage of a CNN prior to its actual application[69].

In this work, we demonstrate how a CNN can be applied to the detection of the GWs from CCSNe and the classification of their explosion mechanisms for two networks of four detectors for a range of distances from 10 kpc to 200 kpc. The first network consists of aLIGO, AdVirgo and KAGRA. For the second network, we also include the detectors of AdVirgo and KAGRA, but replace the two detectors of aLIGO with a modest set of planned upgrade version of them - LIGO A+ in Hanford and Livingston[70, 71]. The remaining of this paper is constructed as follows. In section II, we will present a brief explanation of the concept of CNN as well as the CNN we used for this work. In section III, we discuss the waveforms we used in this work and the procedure with which we generated the data. The results and a discussion will be shown in section IV, followed by a conclusion in section V.

II. CONVOLUTIONAL NEURAL NETWORK

A CNN is a computational processing system that is composed of interconnected layers of computational nodes[72]. The nodes are known as neurons and are associated with an activation function. The activation function can perform an elementwise nonlinear operation to the input of the layer. In a CNN, there are three types of layers: convolutional layer, max-pooling layer, and fully connected layer[72]. Convolutional layer performs the mathematical operation of convolution between the weights of the layer's neurons and the input to that layer. Max-pooling layer is just a down-sampling layer that down samples the input along its dimensionality. It can reduce the computational cost by decreasing the number of parameters of the CNN. Fully connected layer is a layer that connects every neuron in its layer to every neuron of its immediate previous and next layer. In the case of detection and classification, fully connected layers are used to compute class scores.

When multiple layers of these three types are stacked and connected one after the other, a CNN has been formed, where the output of each layer is the input of the next layer. How these layers are connected in a CNN is known as the architecture of the CNN. It describes the structure of a CNN and the number of neurons in each layer. In general, the first layer in a CNN, also known as the input layer, is often a convolutional layer, while the last layer or the output layer is often a fully connected layer with an associated loss function. The architecture of the rest of the CNN should depend on the specific task that the CNN is trained to solve. An over-complicated model with too many trainable parameters is easier to result in overfit and harder to train, while a too-simple CNN will have a hard time capturing the feature inherent to the input. In addition, the number of layers and neurons of each layer in a CNN are known as hyperparameters. Other hyperparameters include the parameters of max-pooling layers, type of activation functions, learning rate of the CNN, and the application of specific deep learning techniques if there is any. The optimal combination of hyperparameters and the architecture is sought by trail and error, and fine tunning.

During the training stage, the weights of the neurons in a CNN are updated using an algorithm called back propagation[73]. The output of the CNN is used as an input to the loss function associated with the output layer. The back propagation algorithm will then compute the gradient of the loss function which is used to adjust the values of the weights of the neurons in each layer and minimise the loss function. When the loss function is minimised, the CNN will classify its input into the correct category with the highest confidence. The process of achieving the minimisation of the loss function during the training stage is referred to as learning.

In this work, we employ a CNN of 8 convolutional layers, 3 max-pooling layers, and 3 fully connected layers. The exact architecture of the CNN is shown in Table I and illustrated in Figure 1. Since the problem we are trying to solve is a problem of multi-class classification, the loss function employed for this work is categorical cross entropy[74], given by

$$L(y, \hat{y}) = -\sum_{j=1}^{M} \sum_{i=1}^{C} (y_{ij} log(\hat{y}_{ij})), \tag{1}$$

TABLE I: The architecture of the CNN

Layer	Type	Neurons	Filter size	Act. Fun
1	Conv	9	64	Elu
2	Max-pool		12	
3	Conv	9	12	Elu
4	Max-pool		6	
5	Conv	11	8	Elu
6	Conv	11	4	Elu
7	Conv	13	4	Elu
8	Conv	13	4	Elu
9	Conv	13	4	Elu
10	Conv	13	4	Elu
11	Max-pool		2	
12	Fully-con	64(50%)		Elu
13	Fully-con	32(50%)		Elu
14	Fully-con	2		Softmax

The architecture of the CNN used in this work for the purpose of distinguishing supernovae signal mechanisms and background noise. In the table, Conv means convolutional neural layer, Max-pool max-pooling layer, and Fully-con fully-connected layers. Neuron and Act. Fun mean the number of neurons and the activation function for the layer respectively. The numbers in the bracket for the fully-connected layers are the number used for drop-out.
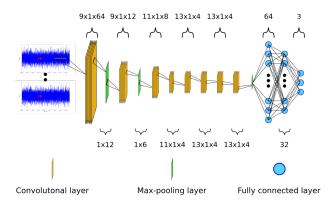

FIG. 1: An illustration of the architecture of the CNN used in this paper for the detection and classification of CCSN signals in noisy data. The CNN consists of 8 convolutional layers, 3 max-pooling layers and 3 fully connected layer including the output layer. The input layer takes the simulated time series of the detectors as input, feeding through the CNN. The CNN will output three probabilities at the last layer. The numbers above or below each layer indicate the kernal size of the layer. For example, the first convolutional layer has 9 filters, each of which is 1 by 64 in size. The elements in the figure are not to scale.

where C is the number of the classes and M is the number of the training samples. For the jth sample and the ith class, y_{ij} is the correponding class value. It is equal to 1 for the true class and 0 otherwise. Similarly, $\hat{y}_{ij}$ is the predicted probability from the CNN for the ith class and the jth sample.

III. DATA

We establish a CNN for the purpose of distinguishing simulated detector time series among three classes, i.e., magnetorotational signals + background noise, neutrino-driven signals + background noise, and pure background noise. For this purpose, it is necessary to prepare training, validation and testing data of these three classes. The training data is used for tuning the weights of the neurons in the layers in the CNN, while validating data is to verify that the CNN is learning the features inherent to the data and testing data is to test the performance of the trained CNN.

In general, the input of a CNN is numerical data. In our case, a data sample is a set of simulated time series stacked together as a $k \times p$ matrix where k is the number of detectors and p the length of the time series. To this end, we use simulated waveforms in the literature. The magnetorotational CCSN signals are taken from[75], where the simulations covered a parameter space of 18 different equations of state and 98 rotation profiles for a progenitor of $12 M_\odot$ generating in total 1824 waveforms. For the neutrino-driven mechanism, we employ 26 waveforms from[37, 76, 77], which cover both 2D simulations with masses of $15 M_\odot$[37], and $12, 15, 20$ and $40 M_\odot$[76], and 3D simulations for a progenitor of $27 M_\odot$[77]. Examples of the simulated waveforms for both mechanisms are shown in Figure 2. Since the waveforms are generated at various distances, sampling rates and durations, it is necessary to normalise the waveforms before they can be used for the generation of the time series. To do this, we first scale the amplitudes of the waveforms by moving the sources to 10 kpc from earth. We then ensure that the sampling rate are identical for all the waveforms by down sampling them to a pre-selected sampling rate (e.g., 4096Hz). The longest duration τ among the waveforms is then identified and each of the remaining waveforms is padded with zeros to this duration. To introduce as less artefact as possible, a high pass filter with a low cut-off frequency equal to 11Hz and a tukey window ($\alpha = 0.08$) are applied prior to the zero padding. To balance the difference in the number of waveforms between the two mechanisms in the final data set, 66 copies of each neutrino-driven waveform are duplicated. After this procedure, the simulated waveforms are then $\mathbf{S}(t) = \{\mathbf{s}_1(t), \mathbf{s}_2(t), ..., \mathbf{s}_m(t)\}$, where m is the number of the waveforms and $\mathbf{s}_q$ is the qth waveform, defined by,

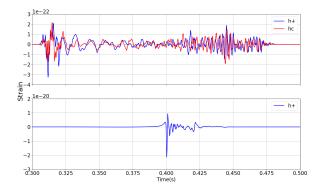$$\mathbf{s}_q(t) = \begin{pmatrix} h_q^+(t) \\ h_q^\times(t) \end{pmatrix}, \tag{2}$$

FIG. 2: Examples of simulated waveforms from both mechanisms used in the work. The top panel shows an example waveform of the neutrino-driven mechanism from[77]. The progenitor is $27M_\odot$ with neutrino heating rate f_{heat} equal to 1.15 in [77]. The $h+$ and the $h\times$ polarisations are shown in blue and red respectively. The bottom panel shows a waveform from the magnetorotational mechanism. The simulation is done assuming a progenitor of $12M_\odot$ with maximum initial rotation rate equal to 12 rad/s and a measure of the degree of differential rotation equal to 300km in Eq.5 in[75]. Only the $h+$ polarisation is shown because the simulation is axis-symmetric and described by only one polarisation. Both sources are assumed to be at 10 kpc from earth.

where $h_q^+(t)$ and $h_q^\times(t)$ are the two polarisations of the waveform and t the time. It is worth noting that some of the waveforms used in the work are generated with the simulations being axis-symmetric. This means for these simulations, the GWs are entirely described by one polarisation $h_q^+(t)$. For these waveforms, the corresponding $h_q^\times(t)$ are a zero vector. In this work, we perform simulations for distances equal to 10, 20, 30, 40, 50, 80, 100, 150 and 200 kpc. If for a training session, the distance d_L interested is not 10 kpc, the following relationship is employed to scale the amplitudes of the waveforms,

$$\mathbf{s}_q^{d_L}(t) = \frac{10\text{kpc} \times \mathbf{s}_q(t)}{d_L}, \qquad (3)$$

where $\mathbf{s}_q^{d_L}(t)$ is the amplitude of the signal if the source is at d_L from earth.

The next step is to generate simulated time series for the GW detectors in a network using $\mathbf{S}(t)$. Since the purpose of building a CNN is to categorise an input data into three exclusive classes, in total three types of time-series are generated. For the time series containing either a magnetorotational signal or a neutrino-driven signal, we start by selecting a waveform $\mathbf{s}_q$ from $\mathbf{S}(t)$. A random location of (right ascension, declination) = (α, δ) in the sky is selected from a uniform distribution on α and a uniform distribution on the sine of δ. The corresponding antenna patterns $\mathbf{F}(\alpha, \delta, t)$, given by the following equation,

tion,

$$\mathbf{F}(\alpha, \delta, t) = \begin{pmatrix} f_1^+(\alpha, \delta, t) & f_1^\times(\alpha, \delta, t) \\ \vdots & \vdots \\ f_k^+(\alpha, \delta, t) & f_k^\times(\alpha, \delta, t) \end{pmatrix}, \qquad (4)$$

is computed, where $f^+(\alpha, \delta, t)$ and $f^\times(\alpha, \delta, t)$ are the antenna pattern functions for the two polarizations and k is the number of detectors as defined above. The relative delays in the arrival times for the location are also computed and applied to the selected waveform. The delay in arrival time between a detector and the center of the earth is given by,

$$\Delta t = \frac{\mathbf{n} \cdot \mathbf{r}}{c}, \qquad (5)$$

where $\mathbf{n}$ is the propagation direction of the GW, c the speed of light, and $\mathbf{r}$ the location vector of the detector relative to the center of the Earth. The resulting signal $\mathbf{h}_j(\alpha, \delta, t)$ as received at the detectors is then given by the following equation,

$$\mathbf{h}_j(\alpha, \delta, t) = \mathbf{F}(\alpha, \delta, t) \times \mathbf{s}_q(t), \qquad (6)$$

where the time delays Δt are absorbed into the corresponding notations and the subscript j is defined in Eq.1. Next, we generate independent Gaussian noise $\mathbf{N}_j(t) = \{\mathbf{n}_{1j}(t), \mathbf{n}_{2j}(t), ..., \mathbf{n}_{kj}(t)\}'$ for each detector in the network using their respective power spectral densities. In the above notation, the symbol $'$ indicates transpose. The duration of the generated noise is 1.7 times longer than $\mathbf{h}_j(\alpha, \delta, t)$. A random number x will then be generated determining where in the generated noise the signal will be placed, as given by,

$$\mathbf{d}_j(\alpha, \delta, t, x) = \begin{cases} \mathbf{N}_j(t) & t < x; \\ \mathbf{N}_j(t) + \mathbf{h}_j(\alpha, \delta, t) & x \le t \le \tau + x; \\ \mathbf{N}_j(t) & t > \tau + x. \end{cases} \qquad (7)$$

This is to avoid the possibility that the CNN learns human artefact instead of common features of the waveforms by having the signals always starting at the same place. For each time series of the background noise class, independently simulated Gaussian background noise of the same duration as that of the other classes are generated for each detectors in the network. This means a data sample for this class is defined as,

$$\mathbf{d}_j(t) = \mathbf{N}_j(t). \qquad (8)$$

Since training a CNN of a given structure to its maximum capacity requires many data samples, we augment our data set by iterating over $\mathbf{S}(t)$ until for each waveform in $\mathbf{S}(t)$, we have 23 data samples $\mathbf{d}$ generated using the procedure described above. This step is essential for the CNN to learn the features of the waveforms and to identify them effectively under different noise scenarios. After that, we generate independent noise realisations for
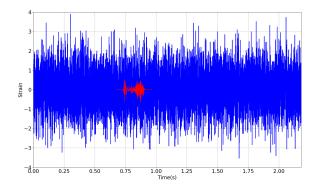
FIG. 3: Representative example of the simulated time series used to train/validate/test the CNN. The blue shows a whitened time series with a signal buried in Gaussian noise of unit variance. The red shows the same signal free of noise and whitened. The signal is the same as the one shown in the top panel in Figure 2 but with antenna pattern applied.

TABLE II: Detector networks

Network	Detectors	Acronym
1	aLIGO H* aLIGO L* AdVirgo KAGRA	HLVK
2	LIGO A+ H LIGO A+ L AdVirgo KAGRA	H+L+VK

The networks of detectors used in this work.
* H refers to the detector in Hanford and L the detector in Livingston.

the background noise class. The entire data set $\mathbf{D}$ defined below, has roughly $l = 1.2 \times 10^6$ data samples, where each class has approximately 4×10^5 data samples.

$$\mathbf{D}(\boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\delta_t}, \mathbf{t}, \mathbf{x}) = \begin{pmatrix} \mathbf{d}_1(\alpha, \delta, \delta_t, t, x) \\ \vdots \\ \mathbf{d}_o(\alpha, \delta, \delta_t, t, x) \\ \vdots \\ \mathbf{d}_l(t) \end{pmatrix} . \qquad (9)$$

In the above equation, the subscript o is the sum of the data samples in which there is a GW signal in the time series. The final step is to whiten $\mathbf{D}$ using the power spectral densities. In Figure 3, a representative example of time series for one detector is shown. A data sample consists of similar time series for all the detectors in the network. The data samples are then split, with 1×10^4 samples being randomly taken for validation, 10% of the rest for testing and 90% for training. When the training is finished for a distance, the above described procedure will be repeated for another distance until the training for all distances have been carried out for a network. The entire procedure is then repeated for another network of GW detectors. As mentioned previously, the CNN will be trained for two networks of GW detectors. We present the networks in Table II. For the remaining of the paper, we will use their acronyms to refer to the networks.

IV. RESULT AND DISCUSSION

After the CNN is trained, we can estimate its performance using the testing samples. The results are presented in this section. One of the most used and convenient ways to determine the classifying performance of a model is to plot the receiver operator characteristic (ROC) curve. A ROC shows the performance of
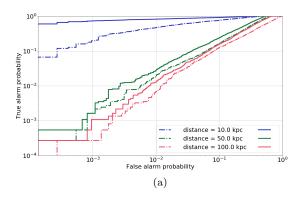
a classifying model by showing the true alarm probability (TAP) at a given false alarm probability (FAP). Since ROC is usually plotted for model distinguishing two classes, for a multi-class classification problem, a ROC for a class should be viewed as the class versus the others. This means in this context, FAP means the fraction of samples from other classes misidentified as a sample from the class the ROC is associated with. TAP is identical to that of two-class classification problem and indicates the fraction of samples correctly identified. For a given FAP, a model with a higher TAP is considered more capable than a model with a lower TAP. In Figure 4, we show the ROCs for both the mechanisms and GW detector networks tested in this work. For simplicity, we show only the results for three distances, namely, 10, 50 and 100 kpc. For all the distances tested, the CNN achieves a higher TAP for any given FAPs for magnetorotational than neutrino-driven signals. That is not surprising as the amplitudes for magnetorotational signals are higher than that of neutrino-driven signals.

We can also show the classification efficiency of the CNN as a function of distance. This is done by fixing the FAP and plotting the TAP. The results for three chosen FAPs are shown in Figure 5. In this figure, a similar trend is seen that magnetorotational signals are easier for the CNN to identify than neutrino-driven signals at any distance. For magnetorotational signals from sources located at 50 kpc at FAP = 0.1, the CNN achieves a TAP of 95% and 85% for the H+L+VK and HLVK respectively. At a more restrict FAP such as FAP= 0.001, the CNN still achieves TAPs 90% and 68% for sources at the same distance for the two networks respectively. Such a range is well into the Large Magellanic Cloud and covers the satellite galaxies in between[78, 79]. For sources at 100 kpc, the TAPs are all larger than 65% for H+L+VK for all chosen FAPs, and even close to 80% if the FAP is 0.1. Even for sources at 150 and 200 kpc, the TAPs are 63% and 48% respectively for the same FAP indicating that with such a GW network, it is possible to detect magnetorotational CCSN signals out to such a distance. On the other hand, it is more difficult for the CNN to detect and
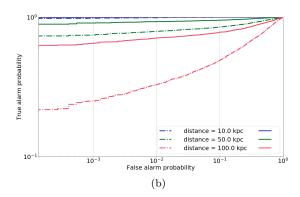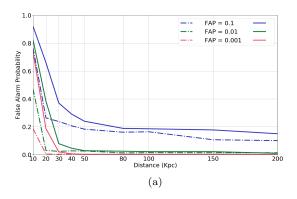
FIG. 4: ROC curves showing the classification performance of the CNN for CCSN signals of different explosion mechanism at three distances. For the distances shown in the both panels, the solid lines in both panels are for signals from the magnetorotational mechanism, while the dashed lines are for the neutrino-driven mechanism. The left panel shows the result for the network of HLVK, while the panel on the right shows that for H+L+VK.
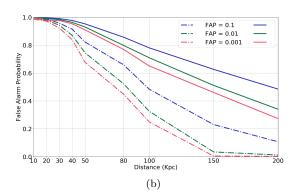

FIG. 5: Efficiency curves showing the classification ability of the CNN as a function of distance for both mechanisms and networks. In both panels, the solid lines show the results for the magnetorotational mechanism, while the dashed lines for the neutrino-driven mechanism. The left panel shows the result for the network of HLVK, while the panel on the right shows that for H+L+VK. Three FAPs are chosen, i.e.: blue for FAP = 0.1, green for FAP = 0.01, red for FAP = 0.001.
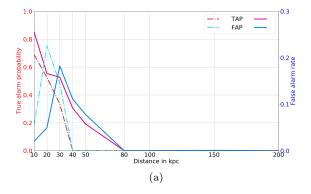
classify the neutrino-driven signals, due to their weaker amplitudes. Nonetheless, for sources at 10 kpc, the CNN achieves TAPs higher than 92% and 76% for H+L+VK and HLVK respectively if the FAP is 0.1. This means that the GW from a Galactic CCSN signal is likely to be detected and classified with either of these networks.

In reality, since the output of a CNN is probabilities indicating how likely the input belongs to each of the classes, it may be desired to set a threshold on which the decision whether the input belong to a class is made. For example, an input will be classified into a class if the corresponding probability is larger than the pre-selected threshold. In such a scenario, the FAP and TAP would be affected by the choice of the threshold. We show such a result in Figure 6 assuming a threshold of 0.5. For H+L+VK and magnetorotational signals at 10 kpc, the TAP is 99%. If the distance is extended to 80 kpc, the TAP is still close to 80%. For the largest distances tested in this work, 150 and 200 kpc, the TAPs are 49% and

34% respectively. For HLVK, the TAP is 98% and 59% for sources at 10 and 80 kpc respectively. The TAPs decrease to 10% and 9% for sources at 150 and 200 kpc. Throughout the distances, the CNN maintains a FAP no larger than 0.04 for both networks. For neutrino-driven signals, it is shown that HLVK has a TAPs of 69% at 10 kpc while it is 85% at 10 kpc for H+L+VK. Both of the networks have FAPs close to or less than 0.2.

V. CONCLUSION

We have demonstrated how to apply a CNN for the purpose of distinguishing GW detector time series among magnetorotational, neutrino-driven and background noise. We trained the CNN using 1.2×10^5 samples of simulated time series of two GW detector networks. The data samples for each classes consisted of approximately 4×10^5 samples.
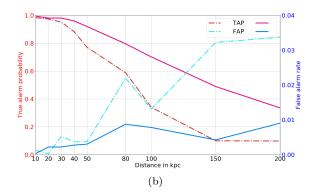
FIG. 6: Efficiency curves showing the ability of the CNN in distinguishing input data with a fixed decision threshold, and their corresponding FAPs. The left panel shows the results for the neutrino-driven mechanism, and the right shows that for the magnetorotational mechanism. In both panels, the solid lines show the TAPs and FAPs for H+L+VK, and the dashed lines show those for HLVK.

We have shown that with a network of HLVK, once trained, a CNN could achieve a TAP of $\geq 85\%$ or $\geq 68\%$ for magnetorotational signals within 50 kpc when the FAP was 0.1 or 0.001 respectively. For the neutrino mechanism, the weaker amplitudes of the waveforms result in a shorter distance. Nonetheless, the trained CNN achieved a TAP of 76% and a FAP of 0.1 for sources at 10 kpc. This indicates sources within the Large Magellanic Cloud are likely to be detectable to the network of HLVK if the explosion mechanism is magnetorotational or a Galactic CCSN event is likely to be detectable if the explosion mechanism is the neutrino-driven mechanism.

Using a network of H+L+VK, we showed that the detection prospective is even more promising. If the explosion is magnetorotational mechanism, the TAP could be higher than 95% for sources with a FAP of 0.1 and 90% with a FAP of 0.001 for sources within 50 kpc. If the distance is extended to 150 or 200 kpc, a TAP of 63% or 48% respectively are still achievable. If the explosion mechanism is the neutrino-driven mechanism, for sources at 10 kpc and a FAP of no less than 0.001, the TAP is at least 73%. These number suggest that with such a network, it is very likely to detect a neutrino-driven supernova event within the Galaxy and a magnetorotational one within the Large Magellanic Cloud and even well beyond such a distance.

ACKNOWLEDGEMENTS

[1] Benjamin P Abbott, Richard Abbott, TD Abbott, et al. Observation of gravitational waves from a binary black hole merger. *Physical review letters*, 116(6):061102, 2016.

[2] Benjamin P Abbott, R Abbott, TD Abbott, et al. Gw151226: observation of gravitational waves from a 22-solar-mass binary black hole coalescence. *Physical review letters*, 116(24):241103, 2016.

[3] Benjamin P Abbott, R Abbott, TD Abbott, et al. Gw170608: Observation of a 19 solar-mass binary black hole coalescence. *The Astrophysical Journal Letters*, 851(2):L35, 2017.

[4] Benjamin P Abbott, R Abbott, TD Abbott, et al. Gw170814: a three-detector observation of gravitational waves from a binary black hole coalescence. *Physical review letters*, 119(14):141101, 2017.

[5] Benjamin P Abbott, Rich Abbott, TD Abbott, et al. Gw170817: observation of gravitational waves from a binary neutron star inspiral. *Physical Review Letters*, 119(16):161101, 2017.

[6] Benjamin P Abbott, R Abbott, TD Abbott, et al. Gravitational waves and gamma-rays from a binary neutron star merger: Gw170817 and grb 170817a. *The Astrophysical Journal Letters*, 848(2):L13, 2017.

[7] Benjamin P Abbott, Richard Abbott, TD Abbott, et al. Multi-messenger observations of a binary neutron star merger. *Astrophys. J. Lett*, 848(2):L12, 2017.

[8] Yoichi Aso, Yuta Michimura, Kentaro Somiya, et al. Interferometer design of the kagra gravitational wave detector. *Physical Review D*, 88(4):043007, 2013.

[9] Kentaro Somiya. Detector configuration of kagra–the japanese cryogenic gravitational-wave detector. *Classical and Quantum Gravity*, 29(12):124007, 2012.

[10] Benjamin P Abbott, R Abbott, TD Abbott, et al. Prospects for observing and localizing gravitational-wave transients with advanced ligo, advanced virgo and kagra. *Living Reviews in Relativity*, 21(1):3, 2018.

[11] Junaid Aasi, BP Abbott, Richard Abbott, et al. Advanced ligo. *Classical and quantum gravity*, 32(7):074001,

2015.

[12] F Acernese, M Agathos, K Agatsuma, et al. Advanced virgo: a second-generation interferometric gravitational wave detector. *Classical and Quantum Gravity*, 32(2):024001, 2014.

[13] SE Gossan, Patrick Sutton, A Stuver, et al. Observing gravitational waves from core-collapse supernovae in the advanced detector era. *Physical Review D*, 93(4):042002, 2016.

[14] Benjamin P Abbott, Richard Abbott, TD Abbott, et al. First targeted search for gravitational-wave bursts from core-collapse supernovae in data of first-generation laser interferometer detectors. *Physical Review D*, 94(10):102001, 2016.

[15] Katsuhiko Sato and Hideyuki Suzuki. Analysis of neutrino burst from the supernova 1987a in the large magellanic cloud. *Physical review letters*, 58(25):2722, 1987.

[16] E Baron and J Cooperstein. The effect of iron core structure on supernovae. *The Astrophysical Journal*, 353:597–611, 1990.

[17] Hans Albrecht Bethe. Supernova mechanisms. *Reviews of Modern Physics*, 62(4):801, 1990.

[18] Evan O'Connor and Christian D Ott. Black hole formation in failing core-collapse supernovae. *The Astrophysical Journal*, 730(2):70, 2011.

[19] Hans A Bethe and James R Wilson. Revival of a stalled supernova shock by neutrino heating. *The Astrophysical Journal*, 295:14–23, 1985.

[20] Hans-Thomas Janka. Explosion mechanisms of core-collapse supernovae. *Annual Review of Nuclear and Particle Science*, 62:407–451, 2012.

[21] Kei Kotake, Kohsuke Sumiyoshi, Shoichi Yamada, et al. Core-collapse supernovae as supercomputing science: A status report toward six-dimensional simulations with exact boltzmann neutrino transport in full general relativity. *Progress of Theoretical and Experimental Physics*, 2012(1), 2012.

[22] Anthony Mezzacappa, Stephen W Bruenn, Eric J Lentz, et al. Two-and three-dimensional multi-physics simulations of core collapse supernovae: A brief status report and summary of results from the" oak ridge" group. *arXiv preprint arXiv:1405.7075*, 2014.

[23] Tomoya Takiwaki, Kei Kotake, and Yudai Suwa. Three-dimensional simulations of rapidly rotating core-collapse supernovae: finding a neutrino-powered explosion aided by non-axisymmetric flows. *Monthly Notices of the Royal Astronomical Society: Letters*, 461(1):L112–L116, 2016.

[24] Alexander Summa, Hans-Thomas Janka, Tobias Melson, and Andreas Marek. Rotation-supported neutrino-driven supernova explosions in three dimensions and the critical luminosity condition. *The Astrophysical Journal*, 852(1):28, 2018.

[25] Stephen W Bruenn, Eric J Lentz, W Raphael Hix, et al. The development of explosions in axisymmetric ab initio core-collapse supernova simulations of 12–25 stars. *The Astrophysical Journal*, 818(2):123, 2016.

[26] H-Th Janka, K Langanke, Andreas Marek, G Martínez-Pinedo, and B Müller. Theory of core-collapse supernovae. *Physics Reports*, 442(1-6):38–74, 2007.

[27] Sean M Couch and Christian D Ott. The role of turbulence in neutrino-driven core-collapse supernova explosions. *The Astrophysical Journal*, 799(1):5, 2015.

[28] John M Blondin, Anthony Mezzacappa, and Christine DeMarino. Stability of standing accretion shocks, with an eye toward core-collapse supernovae. *The Astrophysical Journal*, 584(2):971, 2003.

[29] JM LeBlanc and JR Wilson. A numerical example of the collapse of a rotating magnetized star. *The Astrophysical Journal*, 161:541, 1970.

[30] Adam Burrows, Luc Dessart, Eli Livne, Christian D Ott, and Jeremiah Murphy. Simulations of magnetically driven supernova and hypernova explosions in the context of rapid rotation. *The Astrophysical Journal*, 664(1):416, 2007.

[31] Tomoya Takiwaki, Kei Kotake, and Katsuhiko Sato. Special relativistic simulations of magnetically dominated jets in collapsing massive stars. *The Astrophysical Journal*, 691(2):1360, 2009.

[32] SG Moiseenko, GS Bisnovatyi-Kogan, and NV Ardeljan. A magnetorotational core-collapse model with jets. *Monthly Notices of the Royal Astronomical Society*, 370(1):501–512, 2006.

[33] Philipp Mösta, Sherwood Richers, Christian D Ott, et al. Magnetorotational core-collapse supernovae in three dimensions. *The Astrophysical Journal Letters*, 785(2):L29, 2014.

[34] SE Woosley and Alexander Heger. The progenitor stars of gamma-ray bursts. *The Astrophysical Journal*, 637(2):914, 2006.

[35] S-C Yoon and Norbert Langer. On the evolution of rapidly rotating massive white dwarfs towards supernovae or collapses. *Astronomy & Astrophysics*, 435(3):967–985, 2005.

[36] SE De Mink, N Langer, RG Izzard, Hugues Sana, and Alex de Koter. The rotation rates of massive stars: the role of binary interaction through tides, mass transfer, and mergers. *The Astrophysical Journal*, 764(2):166, 2013.

[37] Christian D Ott. The gravitational-wave signature of core-collapse supernovae. *Classical and Quantum Gravity*, 26(6):063001, 2009.

[38] Kei Kotake. Multiple physical elements to determine the gravitational-wave signatures of core-collapse supernovae. *Comptes Rendus Physique*, 14(4):318–351, 2013.

[39] Kei Kotake, Wakana Iwakami, Naofumi Ohnishi, and Shoichi Yamada. Stochastic nature of gravitational waves from supernova explosions with standing accretion shock instability. *The Astrophysical Journal Letters*, 697(2):L133, 2009.

[40] Ik Siong Heng. Rotating stellar core-collapse waveform decomposition: a principal component analysis approach. *Classical and Quantum Gravity*, 26(10):105005, 2009.

[41] Christian Röver, Marie-Anne Bizouard, Nelson Christensen, et al. Bayesian reconstruction of gravitational wave burst signals from simulations of rotating stellar core collapse and bounce. *Physical Review D*, 80(10):102004, 2009.

[42] Jade Powell, Daniele Trifirò, Elena Cuoco, Ik Siong Heng, and Marco Cavaglià. Classification methods for noise transients in advanced gravitational-wave detectors. *Classical and Quantum Gravity*, 32(21):215012, 2015.

[43] Jade Powell, Alejandro Torres-Forné, Ryan Lynch, et al. Classification methods for noise transients in advanced gravitational-wave detectors ii: performance tests on advanced ligo data. *Classical and Quantum Gravity*, 34(3):034002, 2017.

[44] Sofia Suvorova, Jade Powell, and Andrew Melatos. Re-

constructing gravitational wave core-collapse supernova signals with dynamic time warping. *Physical Review D*, 99(12):123012, 2019.

[45] J Logue, CD Ott, IS Heng, P Kalmus, and JHC Scargill. Inferring core-collapse supernova physics with gravitational waves. *Physical Review D*, 86(4):044023, 2012.

[46] William J Engels, Raymond Frey, and Christian D Ott. Multivariate regression analysis of gravitational waves from rotating core collapse. *Physical Review D*, 90(12):124026, 2014.

[47] TZ Summerscales, Adam Burrows, Lee Samuel Finn, and Christian D Ott. Maximum entropy for gravitational wave data analysis: Inferring the physical parameters of core-collapse supernovae. *The Astrophysical Journal*, 678(2):1142, 2008.

[48] Yekta Gürsel and Massimo Tinto. Near optimal solution to the inverse problem for gravitational-wave bursts. *Physical Review D*, 40(12):3884, 1989.

[49] Malik Rakhmanov. Rank deficiency and tikhonov regularization in the inverse problem for gravitational-wave bursts. *Classical and Quantum Gravity*, 23(19):S673, 2006.

[50] Kazuhiro Hayama, Soumya D Mohanty, Malik Rakhmanov, and Shantanu Desai. Coherent network analysis for triggered gravitational wave burst searches. *Classical and Quantum Gravity*, 24(19):S681, 2007.

[51] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[52] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, et al. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014.

[53] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[54] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*, 2014.

[55] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.

[56] Christian Szegedy, Wei Liu, Yangqing Jia, et al. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[57] Igor Kononenko. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in medicine*, 23(1):89–109, 2001.

[58] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[59] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[60] Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. In *European conference on computer vision*, pages 649–666. Springer, 2016.

[61] Andrej Karpathy and Li Fei-Fei. Deep visual-semantic alignments for generating image descriptions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3128–3137, 2015.

[62] Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. Neural architectures for named entity recognition. *arXiv preprint arXiv:1603.01360*, 2016.

[63] Nikhil Mukund, Sheelu Abraham, Shivaraj Kandhasamy, Sanjit Mitra, and Ninan Sajeeth Philip. Transient classification in ligo data using difference boosting neural network. *Physical Review D*, 95(10):104059, 2017.

[64] Michael Zevin, Scott Coughlin, Sara Bahaadini, et al. Gravity spy: integrating advanced ligo detector characterization, machine learning, and citizen science. *Classical and quantum gravity*, 34(6):064003, 2017.

[65] Daniel George, Hongyu Shen, and EA Huerta. Deep transfer learning: A new deep learning glitch classification method for advanced ligo. *arXiv preprint arXiv:1706.07446*, 2017.

[66] Hunter Gabbard, Michael Williams, Fergus Hayes, and Chris Messenger. Matching matched filtering with deep networks for gravitational-wave astronomy. *Physical review letters*, 120(14):141103, 2018.

[67] Daniel George and EA Huerta. Deep learning for real-time gravitational wave detection and parameter estimation: Results with advanced ligo data. *Physics Letters B*, 778:64–70, 2018.

[68] P Astone, P Cerdá-Durán, I Di Palma, et al. New method to observe gravitational waves emitted by core collapse supernovae. *Physical Review D*, 98(12):122002, 2018.

[69] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[70] John Miller, Lisa Barsotti, Salvatore Vitale, et al. Prospects for doubling the range of advanced ligo. *Physical Review D*, 91(6):062005, 2015.

[71] B. Lantz, S. Danilishin, S. Hild, et al. Ligo instrument science white paper. 2018.

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

[73] Yann LeCun, D Touresky, G Hinton, and T Sejnowski. A theoretical framework for back-propagation. In *Proceedings of the 1988 connectionist models summer school*, volume 1, pages 21–28. CMU, Pittsburgh, Pa: Morgan Kaufmann, 1988.

[74] Martín Abadi, Ashish Agarwal, Paul Barham, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.

[75] Sherwood Richers, Christian D Ott, Ernazar Abdikamalov, Evan OConnor, and Chris Sullivan. Equation of state effects on gravitational waves from rotating core collapse. *Physical Review D*, 95(6):063019, 2017.

[76] Jeremiah W Murphy, Christian D Ott, and Adam Burrows. A model for gravitational wave emission from neutrino-driven core-collapse supernovae. *The Astrophysical Journal*, 707(2):1173, 2009.

[77] Christian D Ott, Ernazar Abdikamalov, Philipp Mösta, et al. General-relativistic simulations of three-dimensional core-collapse supernovae. *The Astrophysical Journal*, 768(2):115, 2013.

[78] Igor D Karachentsev, Valentina E Karachentseva, Walter K Huchtmeier, and Dmitry I Makarov. A catalog of neighboring galaxies. *The Astronomical Journal*, 127(4):2031, 2004.

[79] V Belokurov, Daniel B Zucker, NW Evans, et al. Cats and dogs, hair and a hero: a quintet of new milky way companions. *The Astrophysical Journal*, 654(2):897, 2007.