engineering proceedings

Proceeding Paper

Case Study of Binary Hypothesis Test Using ML †

Shang-Hua Chin [1,*] and Cheng-Yu Chin [2]

[1] Department of Finance, National Central University, Taoyuan 320, Taiwan
[2] Department of Information Engineering and Computer Science, Feng Chia University, Taichung 407, Taiwan; cheneychin0723@gmail.com
* Correspondence: shanghuachin@gmail.com or d1149621@o365.fcu.edu.tw
† Presented at the 2024 Cross Strait Conference on Social Sciences and Intelligence Management, Shanghai, China, 13–15 December 2024.

Abstract

Artificial intelligence has attracted much attention due to its learning capability to solve versatile problems. Using a convolutional neural network in machine learning (ML), we investigated the binary hypothesis test, which is a fundamental problem in management and business. The simulation results showed that the proposed method is comparable with the conventional optimum likelihood ratio test for the aspect of type I and II errors. Moreover, the learning capability of ML is promising for complicated data, the properties of which, such as probability distribution and/or statistical data, i.e., mean, variance, and others, are not known.

Keywords: binary hypothesis test; convolutional neural network; likelihood ratio test; ML

1. Introduction

The binary hypothesis test is used to select better hypotheses between $\mathcal{H}_0$ and $\mathcal{H}_1$ infers based on measurement data. Many applications are used in management and business [1,2], communications [3,4], and biology [5–7]. The hypothesis test enables the analysis and evaluation of the causes and effects of important management decisions. Using the hypothesis test, data-driven decision-makers translate data into their business strategy.

Traditionally, the hypothesis test is conducted using the Bayesian approach. Recently, machine learning (ML) has attracted considerable attention due to its outstanding classification capability using features embedded in observed data. Characterized by the use of data and algorithms, ML is used to solve problems by imitating the way that humans learn and improve their performance (or intelligence). Convolutional neural networks (CNNs) [8,9] are popular deep learning technologies that are often used in image or speech recognition, object detection, and other fields. Compared with the traditional multilayer perceptron (MLP), the CNN introduces the concept of windows. This allows the CNN to effectively capture high-level features of the data observed through these windows. CNNs usually contain various layers, such as convolutional layers, activation functions, pooling layers, and fully connected layers [8].

In contrast to the conventional likelihood ratio test, we studied the binary hypothesis test using the ML approach. Using the CNN, a solution was proposed, and the results were compared with that of the optimum likelihood ratio test in this study. Consequently, the learning capability of CNN models is promising for solving the considered problem of binary hypothesis tests in wide applications.

check for updates

Academic Editors: Teen-Hang Meen, Liza Lee and Cheng-Fu Yang

Published: 14 July 2025

Citation: Chin, S.-H.; Chin, C.-Y. Case Study of Binary Hypothesis Test Using ML. *Eng. Proc.* **2025**, *98*, 37. https://doi.org/10.3390/engproc2025098037

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

The rest of this article is organized as follows. Section 2 introduces the problem. Section 3 presents two case studies employing the conventional likelihood ratio test. Section 4 introduces the proposed CNN scheme, and Section 5 demonstrates simulation results and concludes this study.

2. Problem Formulation

Without a loss of generality, the binary hypothesis test was investigated, assuming that data were sampled from the stationary Gaussian random process. Therefore, the joint probability density function (PDF) of N data samples $\mathbf{x} = [x_0, x_1 \ldots x_{N-1}]^T$ under hypothesis $\mathcal{H}_i$, $I = 0, 1$, was expressed as follows:

$$f(\mathbf{x}|\mathcal{H}_i) = \prod_{n=0}^{N-1} \frac{\exp\left(-\frac{1}{2}\left(\frac{x_n-\mu_i}{\sigma_i}\right)^2\right)}{\sigma_i\sqrt{2\pi}} \tag{1}$$

where x_n denotes data samples; $n = 0, 1, \ldots, N-1$; $\prod$ denotes the product operation; μ_i and σ_i denote the mean and standard deviation of x_n under hypothesis $\mathcal{H}_i$, respectively; and $\mu_1 > \mu_0$. Given the data samples, the true hypothesis must be selected. That is, based on $\mathbf{x}$ and its joint PDF(1), the most probable hypothesis is inferred as follows:

$$\begin{cases} \mathcal{H}_0 : \mathbf{x} \sim f(\mathbf{x}|\mathcal{H}_0) \\ \mathcal{H}_1 : \mathbf{x} \sim f(\mathbf{x}|\mathcal{H}_1) \end{cases}. \tag{2}$$

3. Case Study Using Likelihood Ratio Test

We explored two cases using the likelihood ratio test.

3.1. Case A: Different Mean with Same Variance

Since mean and variance are basic statistics of Gaussian random variables, their effects on the considered problem were studied. First, we investigated a simple condition with a different mean and the same variance, i.e., $\mu_0 \neq \mu_1$ ($\mu_1 > \mu_0$) and $\sigma_0^2 = \sigma_1^2$. According to the likelihood ratio test,

$$f(\mathrm{x}|\mathcal{H}_1) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} f(\mathrm{x}|\mathcal{H}_0) \tag{3}$$

After several manipulations, Equation (4) was used.

$$\frac{1}{N}\sum_{n=0}^{N-1} x_n \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta_A \tag{4}$$

where $\eta_A = \frac{\mu_0+\mu_1}{2}$.

The decision metrics (Equation (4)) are comprehensive. When the average value of $\boldsymbol{x}$ is less than η_A (or the average value of means μ_0 and μ_1), $\mathcal{H}_0$ is decided; otherwise, $\mathcal{H}_1$ is inferred.

Under $\mathcal{H}_0$ (i.e., $x_n \sim \aleph(\mu_0, \sigma_0^2)$, $\forall n$), let $y_A \equiv \frac{1}{N}\sum_{n=0}^{N-1} x_n$; one has $y_A \sim \aleph\left(\mu_{y_A}, \sigma_{y_A}^2\right)$, where $\aleph(\cdot)$ denotes the PDF of a Gaussian random variable, $\mu_{y_A} = \mu_0$ is the mean of y_A,

and $\sigma_{y_A}^2 = \sigma_0^2/N$ is the variance of y_A. Therefore, the type I error of case A, α_A, can be subsequently written as follows:

$$\begin{aligned} \alpha_A &= \text{Prob}(y_A > \eta_A) \\ &= \int_{\eta_A}^{\infty} \frac{\exp\left(-\frac{1}{2}\left(\frac{y_A-\mu_{y_A}}{\sigma_{y_A}}\right)^2\right)}{\sigma_{y_A}\sqrt{2\pi}} dy_A \end{aligned} \tag{5}$$

where $\text{Prob}(\cdot)$ denotes the probability.

Apparently, α_A does not have the closed-form expression. However, since $z_A = \frac{y_A-\mu_{y_A}}{\sigma_{y_A}}$ has the standard normal distribution, α_A is rewritten as follows:

$$\begin{aligned} \alpha_A &= \text{Prob}\left(\frac{y_A-\mu_{y_A}}{\sigma_{y_A}} > \frac{\eta_A-\mu_{y_A}}{\sigma_{y_A}}\right) \\ &= Q\left(\frac{\eta_A-\mu_{y_A}}{\sigma_{y_A}}\right), \end{aligned} \tag{6}$$

where $Q(\cdot)$ denotes the Q function, i.e.,

$$Q(x) = \int_x^{\infty} \frac{\exp\left(-\frac{1}{2}u^2\right)}{\sqrt{2\pi}} du. \tag{7}$$

Under $\mathcal{H}_1$ (i.e., $x_n \sim \aleph\left(\mu_1, \sigma_1^2\right), \forall n$), let $y_A \equiv \frac{1}{N}\sum_{n=0}^{N-1} x_n$; one has $y_A \sim \aleph\left(\mu_{y_A}, \sigma_{y_A}^2\right)$, where $\aleph(\cdot)$ denotes the PDF of a Gaussian random variable, $\mu_{y_A} = \mu_1$ is the mean of y_A, and $\sigma_{y_A}^2 = \sigma_1^2/N$ is the variance of y_A. Therefore, the type II error of case A, β_A, can be subsequently written as follows:

$$\begin{aligned} \beta_A &= \text{Prob}(y_A < \eta_A) \\ &= \int_{-\infty}^{\frac{\eta_A-\mu_{y_A}}{\sigma_{y_A}}} \frac{\exp\left(-\frac{1}{2}z_A^2\right)}{\sqrt{2\pi}} dz_A \end{aligned} \tag{8}$$

where $z_A = \frac{y_A-\mu_{y_A}}{\sigma_{y_A}}$. Let $w_A = -z_A$:

$$\begin{aligned} \beta_A &= \int_{-\frac{\eta_A-\mu_{y_A}}{\sigma_{y_A}}}^{\infty} \frac{\exp\left(-\frac{1}{2}w_A^2\right)}{\sqrt{2\pi}} dw_A \\ &= Q\left(-\frac{\eta_A-\mu_{y_A}}{\sigma_{y_A}}\right). \end{aligned} \tag{9}$$

3.2. Case B: Different Mean with Different Variance

In this general case, $\mu_0 \neq \mu_1$ ($\mu_1 > \mu_0$) and $\sigma_0^2 \neq \sigma_1^2$. According to the likelihood ratio test (Equation (3)), after several manipulations, the following is derived:

$$\sum_{n=0}^{N-1}\left(\frac{x_n-\mu_0}{\sigma_0}\right)^2 - \sum_{n=0}^{N-1}\left(\frac{x_n-\mu_1}{\sigma_1}\right)^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta_B \tag{10}$$

where $\eta_B = 2N\ln\frac{\sigma_1}{\sigma_0}$ and $\ln(\cdot)$ denote the natural logarithm function.

Even though the decision metric (Equation (10)) is not concise, it is enough to be used as the algorithm of the likelihood ratio test for case B. Since the exact distributions of the decision metric under both hypotheses are involved, we derived their asymptotical expressions when N was large. Hence, under $\mathcal{H}_0$ (i.e., $x_n \sim \aleph\left(\mu_0, \sigma_0^2\right), \forall n$), let $y_B \equiv \sum_{n=0}^{N-1}\left(\frac{x_n-\mu_0}{\sigma_0}\right)^2$ and $w_B \equiv \sum_{n=0}^{N-1}\left(\frac{x_n-\mu_1}{\sigma_1}\right)^2$; according to the central limit theo-

rem, one has $y_B \sim \aleph\left(\mu_{y_B}, \sigma_{y_B}^2\right)$ and $w_B \sim \aleph\left(\mu_{w_B}, \sigma_{w_B}^2\right)$, where $\mu_{y_B} = N$, $\sigma_{y_B}^2 = 2N$, $\mu_{w_B} = N\frac{\sigma_0^2}{\sigma_1^2} + N\frac{(\mu_0-\mu_1)^2}{\sigma_1^2}$, $\sigma_{w_B}^2 = 2N\frac{\sigma_0^4}{\sigma_1^4} + 4N(\mu_0-\mu_1)^2\frac{\sigma_0^2}{\sigma_1^4}$ (Appendix A).

Let $z_B = y_B - w_B$, then $z_B \sim \aleph\left(\mu_{z_B}, \sigma_{z_B}^2\right)$, where $\mu_{z_B} = N\left(1 - \frac{\sigma_0^2}{\sigma_1^2}\right) - N\frac{(\mu_0-\mu_1)^2}{\sigma_1^2}$ and $\sigma_{z_B}^2 = 2N + 2N\frac{\sigma_0^4}{\sigma_1^4} + 4N(\mu_0-\mu_1)^2\frac{\sigma_0^2}{\sigma_1^4} - 4N\frac{\sigma_0^2}{\sigma_1^2}$ (Appendix B). Therefore, the type I error of case B, α_B, is written as follows:

$$\begin{aligned} \alpha_B &= \text{Prob}(z_B > \eta_B) \\ &= Q\left(\frac{\eta_B - \mu_{z_B}}{\sigma_{z_B}}\right). \end{aligned} \tag{11}$$

Under $\mathcal{H}_1$ (i.e., $x_n \sim \aleph\left(\mu_1, \sigma_1^2\right)$, $\forall n$), the statistics of $y_B = \sum_{n=0}^{N-1}\left(\frac{x_n-\mu_0}{\sigma_0}\right)^2$ and $w_B \equiv \sum_{n=0}^{N-1}\left(\frac{x_n-\mu_1}{\sigma_1}\right)^2$ are similar to $y_B \sim \aleph\left(\mu_{y_B}, \sigma_{y_B}^2\right)$ and $w_B \sim \aleph\left(\mu_{w_B}, \sigma_{w_B}^2\right)$, where $\mu_{y_B} = N\frac{\sigma_1^2}{\sigma_0^2} + N\frac{(\mu_1-\mu_0)^2}{\sigma_0^2}$, $\sigma_{y_B}^2 = 2N\frac{\sigma_1^4}{\sigma_0^4} + 4N(\mu_1-\mu_0)^2\frac{\sigma_1^2}{\sigma_0^4}$, $\mu_{w_B} = N$, and $\sigma_{w_B}^2 = 2N$. Additionally, let $z_B = y_B - w_B$, then $z_B \sim \aleph\left(\mu_{z_B}, \sigma_{z_B}^2\right)$, where $\mu_{z_B} = N\left(\frac{\sigma_1^2}{\sigma_0^2} - 1\right) + N\frac{(\mu_1-\mu_0)^2}{\sigma_0^2}$ and $\sigma_{z_B}^2 = 2N + 2N\frac{\sigma_1^4}{\sigma_0^4} + 4N(\mu_1-\mu_0)^2\frac{\sigma_1^2}{\sigma_0^4} - 4N\frac{\sigma_1^2}{\sigma_0^2}$. Therefore, the type II error of case B, β_B, can be written as follows:

$$\beta_B = \text{Prob}(z_B < \eta_B) = Q\left(-\frac{\eta_B - \mu_{z_B}}{\sigma_{z_B}}\right). \tag{12}$$

4. Proposed ML Based on CNN

Table 1 presents the proposed CNN structure for the considered hypothesis test with N = 40. The stride and activation functions of the CNN are one and ReLU, respectively. Compared with the conventional optimum likelihood ratio test, the probability distribution and/or statistical properties of the data samples are not required by CNNs. However, training data and their correct classifications (or labels) are necessary for the ML.

Table 1. CNN-based binary hypothesis test with N = 40.

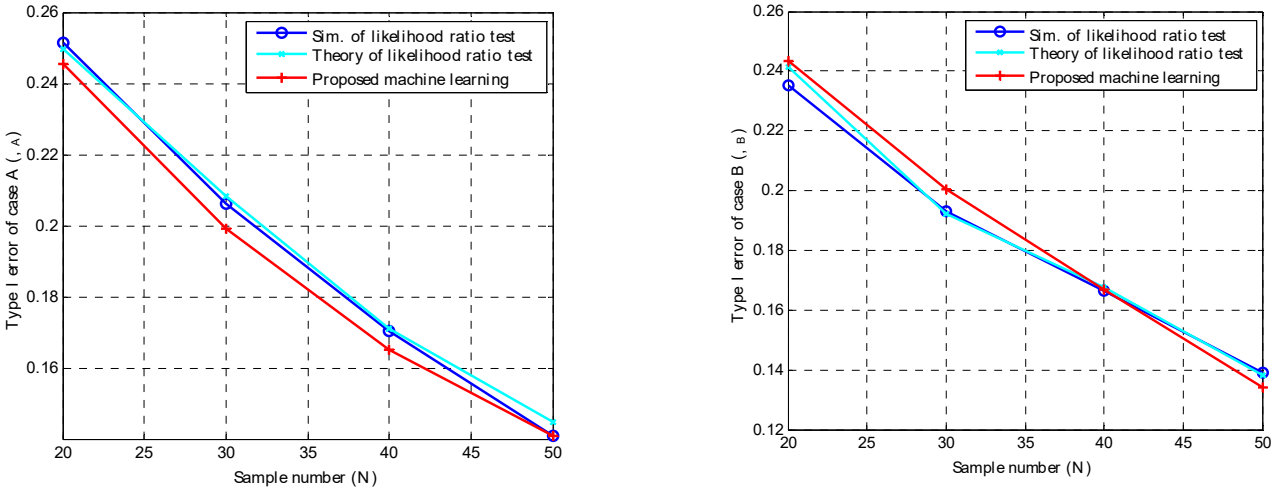
Layer	Output Dimensions	Parameter Number
Input	(40, 1)	0
Convolution	(40, 50)	950 37,550 30,050 22,550 15,050 7550
Fully connected	64	128,064
	32	2080
	16	528
	8	136
	1	9

5. Discussion and Conclusions

Monte Carlo simulations were used to assess the performance of binary hypothesis tests using different approaches in this study. For case A, 20,000 sets of input data were evenly sampled from the Gaussian random variables with $\left(\mu_0,\ \sigma_0^2\right) = (-0.15, 1)$ and $\left(\mu_1,\ \sigma_1^2\right) = (0.15, 1)$ for $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Each set had N data samples. On the contrary, for case B, 20,000 sets of input data were also evenly sampled from the Gaussian

random variables with $(\mu_0, \sigma_0^2) = (-0.15, 1)$ and $(\mu_1, \sigma_1^2) = (0.15, 1.1)$ for $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. In each case, we divided all sample data into two datasets for the proposed CNN: training data (80%) and testing data (20%).

Figure 1a,b present the type I error plotted against the sample number, N, of the proposed ML based on the CNN, as well as the theoretical and simulation results of the likelihood ratio test for cases A and B, respectively. All performances produced similar results. This confirms that the proposed CNN performed similarly to the conventional optimum likelihood ratio test. Notably, similar results were observed for the type II error, which is omitted here.



(**a**) (**b**)

Figure 1. (**a**) Type I error of case A under $\mathcal{H}_0$, i.e., α_A, for the proposed ML based on the CNN and the theoretical and simulation results of the likelihood ratio test; (**b**) type I error of case B under $\mathcal{H}_0$, i.e., α_B, for the proposed ML based on the CNN and the theoretical and simulation results of the likelihood ratio test.

This trial confirmed that the proposed CNN can be adopted for binary decision-making. Accompanied by the sampling method, the learning capability of ML is appropriate for the classification problem with an unknown probability distribution and statistical data, i.e., the mean and variance. By integrating deep learning technology, ML is used for common and/or complicated data, which do not require any confirmed statistical properties.

Author Contributions: Conceptualization, S.-H.C.; methodology, S.-H.C. and C.-Y.C.; validation, C.-Y.C.; formal analysis, S.-H.C.; writing, S.-H.C. and C.-Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All research data were obtained through MATLAB R2012a.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Derivations of $\mu_{y_B}, \sigma_{y_B}^2, \mu_{w_B}$, and $\sigma_{w_B}^2$ under $\mathcal{H}_0$:
By definition,

$$\mu_{y_B} = E[y_B] = \sum_{n=0}^{N-1} \frac{E\left[(x_n - \mu_0)^2\right]}{\sigma_0^2} = N, \tag{13}$$

where $E[\cdot]$ denotes the expectation operator. In addition,

$$\begin{aligned} E\left[y_B^2\right] &= E\left[\sum_{n=0}^{N-1}\frac{(x_n-\mu_0)^2}{\sigma_0^2}\sum_{m=0}^{N-1}\frac{(x_m-\mu_0)^2}{\sigma_0^2}\right] \\ &= \frac{1}{\sigma_0^4}\sum_{n=0}^{N-1}E\left[(x_n-\mu_0)^4\right]+\frac{1}{\sigma_0^4}\sum_{n=0}^{N-1}\sum_{\substack{m=0\\m\neq n}}^{N-1}E\left[(x_n-\mu_0)^2\right]E\left[(x_m-\mu_0)^2\right]. \end{aligned} \tag{14}$$

Remember that

$$E\left[(x_n-\mu_0)^p\right]=\begin{cases}0, & \text{if } p \text{ is odd}\\ \sigma_0^p\prod_{k=1}^{p/2}(2k-1), & \text{if } p \text{ is even}\end{cases}. \tag{15}$$

Hence,

$$E\left[y_B^2\right]=N(N+2). \tag{16}$$

Therefore, using Equations (13) and (16),

$$\sigma_{y_B}^2=E\left[y_B^2\right]-\mu_{y_B}^2=2N. \tag{17}$$

Similarly, we obtained

$$\mu_{w_B}=E[w_B]=\sum_{n=0}^{N-1}\frac{E\left[(x_n-\mu_1)^2\right]}{\sigma_1^2}=N\frac{\sigma_0^2}{\sigma_1^2}+N\frac{(\mu_0-\mu_1)^2}{\sigma_1^2} \tag{18}$$

and

$$\sigma_{w_B}^2=E\left[w_B^2\right]-\mu_{w_B}^2=2N\frac{\sigma_0^4}{\sigma_1^4}+4N(\mu_0-\mu_1)^2\frac{\sigma_0^2}{\sigma_1^4}. \tag{19}$$

Appendix B

Derivation of $\sigma_{z_B}^2$:

Since $z_B=y_B-w_B$ and y_B and w_B are correlated,

$$\sigma_{z_B}^2=\sigma_{y_B}^2+\sigma_{w_B}^2-2\mathrm{Cov}(y_B,w_B) \tag{20}$$

where

$$\mathrm{Cov}(y_B,w_B)=E[y_Bw_B]-\mu_{y_B}\mu_{w_B}. \tag{21}$$

Additionally,

$$\begin{aligned} E[y_Bw_B] &= \sum_{n=0}^{N-1}\sum_{m=0}^{N-1}E\left[\left(\frac{x_n-\mu_0}{\sigma_0}\right)^2\right]E\left[\left(\frac{x_m-\mu_1}{\sigma_1}\right)^2\right] \\ &= N(N+2)\frac{\sigma_0^2}{\sigma_1^2}+N^2(\mu_0-\mu_1)^2\frac{1}{\sigma_1^2}. \end{aligned} \tag{22}$$

Finally, using Equations (13), (17)–(19), (20)–(22), one can determine the following:

$$\sigma_{z_B}^2=2N+2N\frac{\sigma_0^4}{\sigma_1^4}+4N(\mu_0-\mu_1)^2\frac{\sigma_0^2}{\sigma_1^4}-4N\frac{\sigma_0^2}{\sigma_1^2}. \tag{23}$$

References

1. Dayanik, S.; Sezer, S. Multisource Bayesian sequential binary hypothesis testing problem. *Ann. Oper. Res.* **2012**, *201*, 99–130. [CrossRef]

2. Mehdi, I.K.E. An examination of the naive-investor hypothesis in accruals mispricing in Tunisian firms. *J. Int. Financ. Manag. Account.* **2011**, *22*, 131–164. [CrossRef]
3. Chin, W.L. On the noise uncertainty for the energy detection of OFDM signals. *IEEE Trans. Veh. Technol.* **2019**, *68*, 7593–7602. [CrossRef]
4. Chin, W.L.; Le, T.N.; Tseng, C.L. Authentication scheme for mobile OFDM based on security information technology of physical layer over time-variant and multipath fading channels. *Inf. Sci.* **2015**, *321*, 238–249. [CrossRef]
5. Read, J.; Sharma, S. Hypothesis-driven science in large-scale studies: The case of GWAS. *Biol. Philos.* **2021**, *36*, 46. [CrossRef]
6. Li, J.J.; Tong, X. Statistical hypothesis testing versus ML binary classification: Distinctions and guidelines. *Patterns* **2020**, *1*, 100115. [CrossRef] [PubMed]
7. Wei, Z.; Yang, A.; Rocha, L.; Miranda, M.F.; Nathoo, F.S. A Review of Bayesian Hypothesis Testing and Its Practical Implementations. *Entropy* **2022**, *24*, 161. [CrossRef] [PubMed]
8. Chinga, A.; Bendezu, W.; Angulo, A. Comparative Study of CNN Architectures for Brain Tumor Classification Using MRI: Exploring GradCAM for Visualizing CNN Focus. *Eng. Proc.* **2025**, *83*, 22. [CrossRef]
9. Chin, W.L.; Lai, S.C.; Lin, S.W.; Chen, H.H. Pipelined neural network assisted mobility speed estimation over doubly-selective fading channels. *IEEE Wirel. Commun.* **2022**, *31*, 163–168. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.