

TroyNet: A Region-Specific Deep Learning Model for Seismic Phase Picking in Noisy Networks of Northwestern Türkiye

This Accepted Manuscript (AM) is a PDF file of the manuscript accepted for publication after peer review, when applicable, but does not reflect post-acceptance improvements, or any corrections. Use of this AM is subject to the publisher's embargo period and AM terms of use. Under no circumstances may this AM be shared or distributed under a Creative Commons or other form of open access license, nor may it be reformatted or enhanced, whether by the Author or third parties. By using this AM (for example, by accessing or downloading) you agree to abide by Springer Nature's terms of use for AM versions of subscription articles: <https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms>

The Version of Record (VOR) of this article, as published and maintained by the publisher, is available online at: <https://doi.org/10.1007/s00024-026-03938-9>. The VOR is the version of the article after copy-editing and typesetting, and connected to open research data, open protocols, and open code where available. Any supplementary information can be found on the journal website, connected to the VOR.

For research integrity purposes it is best practice to cite the published Version of Record (VOR), where available (for example, see ICMJE's guidelines on overlapping publications). Where users do not have access to the VOR, any citation must clearly indicate that the reference is to an Accepted Manuscript (AM) version.

TroyNet: A Region-Specific Deep Learning Model for Seismic Phase Picking in Noisy Networks of Northwestern Türkiye

Utku UNAL¹, Tolga BEKLER^{1,2,3}

¹Geophysical Engineering, School of Graduate Studies, Canakkale Onsekiz Mart University, Canakkale, 17020, Türkiye.

²Geophysical Engineering, Faculty of Engineering, Canakkale Onsekiz Mart University, Canakkale, 17020, Türkiye.

³Earthquake Research and Application Center, Canakkale Onsekiz Mart University, Canakkale, 17020, Türkiye.

Corresponding author: Utku Unal (23414961001@ogr.comu.edu.tr)

Abstract

Northwestern Türkiye is a region with high earthquake activity under the complex compressional and extensional tectonic regime of the Western Anatolian block between the Eurasian and African plates, with dominant control from the North Anatolian Fault Zone. In seismically active regions, accurate and timely P- and S-wave arrival picking is crucial for earthquake monitoring and parameter estimation. Although several deep learning algorithms have been proposed for this purpose, models trained in one region often degrade when applied to other regions, where the noise conditions and seismic network characteristics are different. In this study, we present TroyNet, a PhaseNet-derived seismic phase-picking architecture tailored for northwestern Türkiye. TroyNet features a U-Net-like architecture and consists of a dilated residual encoder, minimal skip connectivity, nearest-neighbor upsampling, and a positional-transformer bottleneck to capture distant temporal dependencies. For coping with extremely imbalanced classes, our network incorporates a customized class-weighted loss function during training. Noise-based clipping and SNR-based sample weighting were evaluated in controlled ablations; the final configuration excludes sample weighting. TroyNet was compared with a retrained PhaseNet baseline implemented and evaluated under identical conditions. TroyNet achieved higher P- and S-wave recalls with similar level of phase detection precision. These results demonstrate that incorporating region-specific data characteristics, architectural improvements, and tailored training can improve phase picking in noisy environments.

Keywords: TroyNet, Seismic phase picking, machine learning, U-Net, transformer, Northwestern Türkiye

Acknowledgments

- Both authors thank KOERI (Kandilli Observatory and Earthquake Research Center, Bogazici University, Istanbul, Türkiye) for providing public seismograms as additional data for the earthquakes that are subject to this study.
- The computational work in this study was supported by in-kind access to computer facilities and hardware funded by the Çanakkale Onsekiz Mart University Scientific Research Projects (project code: FBA-2024-4926). No direct funding was received.

1. Introduction

Accurate P-wave and S-wave arrival time picking is a fundamental issue in earthquake detection and location computations. Conventional statistical methods such as STA/LTA or Akaike Information Criterion (AIC) pickers are used for automatic phase picking detection, but their sensitivity declines in low-signal-to-noise (SNR) environments or in regions with heterogeneous crustal structure. In contrast, machine learning algorithms have achieved excellent performance in seismic phase picking with better precision and consistency by learning from enormous, labeled datasets like EQTransformer (Mousavi et al., 2020), GPD (Ross et al., 2018), EQCCT (Saad et al., 2023) and PhaseNet (Zhu & Beroza, 2019). Recent results from region-specific and multi-region investigations have confirmed that DL pickers can provide detailed quake data sets far beyond what traditional methods can provide in areas like volcano arcs, rift valleys, and induced seismicity areas (Isken et al., 2025; Büyükakpınar et al., 2025; Heuel et al., 2025; Caredda et al., 2025; Jamalreyhani et al., 2025). Deep neural networks can utilize large training datasets of seismic waveforms to pick weak onsets of signals that are not detectable via conventional techniques: Deep learning pickers have improved the number of detectable seismic phases by including those from newly identified events and from distant, low signal-to-noise ratio (SNR) arrivals, thus increasing the completeness of earthquake catalogs (Aguilar Suarez & Beroza, 2025). Recently, graph-based networks trained on multi-region data have reported higher recall and F1 over PhaseNet & EQTransformer, especially for noisy S-waves (Xu et al., 2024).

Even with these improvements, model deployment to new areas faces the challenge of variability in local seismic characteristics and background noise. A model trained on data for one area would not necessarily extrapolate well to another area without calibration because wave propagation differences, station instrumentation, and noise level could compromise performance. Earlier research established the necessity for domain-specific training or transfer learning to bridge these gaps (Münchmeyer et al., 2022).

The selected region, encompassing the Biga Peninsula in northwestern Türkiye, is characterized by a network of active faults, including the southern branch of the North Anatolian Fault. This area is considered one of Türkiye's most seismically active regions, exhibiting frequent microseismicity and complex interactions among faults (Figure 1). The area exhibits significant crustal heterogeneity and is dominated by active strike-slip faulting along the southern branch of the North Anatolian Fault Zone (Bekler & Gürbüz, 2008). Additionally, the Biga Peninsula and nearby areas have a high level of seismic risk, some parts of the region that extend westward from the main North Anatolian Fault Zone, especially around the Saros–Gaziköy and Yenice–Gönen faults, display patterns that are often linked to large earthquakes (Akol & Bekler, 2013). Recent seismological and geodetic analyses further indicate that the Marmara Sea basin and northwestern Anatolia next to it form a structurally complex and actively deforming region. It is characterized by distributed strain, segmented fault systems, and pronounced lateral variations in crustal properties (Karabulut et al., 2024). These factors underscore the need for automated, accurate and reliable phase picking tools such as TroyNet, which are specifically designed to work well in this challenging, high-risk environment.

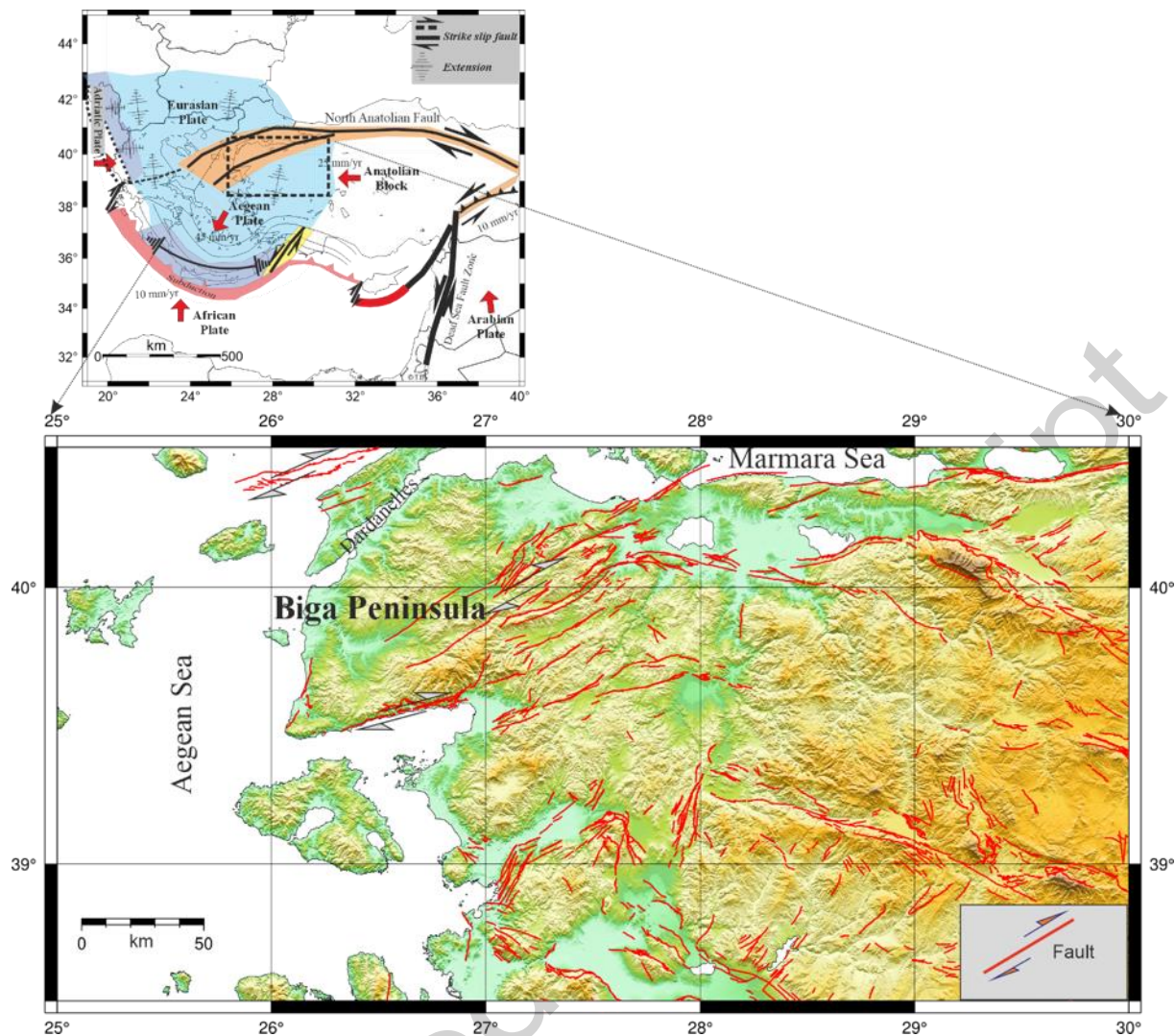


Figure 1. Tectonic setting of northwestern Türkiye and the location of the study area. The upper panel presents the broader tectonic framework of Anatolia and the surrounding plates, while the lower panel focuses on the Biga Peninsula and nearby regions, showing the distribution of mapped active faults. The study area is defined by latitudes 38.60° – 40.65° N and longitudes 25.00° – 29.65° E.

Previous works applied the original PhaseNet (Zhu & Beroza, 2019), which was trained on the Californian dataset, to this region and surrounding areas and reported that it outperformed traditional pickers, yielding more accurate picks and thus improving earthquake location estimates (Unal & Bekler, 2025). While several recent models use different strategies to capture long-range dependencies, such as EQTransformer (Mousavi et al., 2020) which combines convolutional layers, recurrent units, and attention mechanisms or EQCCT (Saad et al., 2023) which employ deep transformer stacks, TroyNet adopts a focused design: it preserves the strengths of the encoder–decoder architecture while introducing only the targeted components: its key modification is a single transformer bottleneck, combined with parameter-free sinusoidal positional encoding that is inserted between the dilated-residual encoder and the decoder. This attention layer allows the network to capture longer-range temporal relationships without introducing depth or complexity of full transformer architectures. TroyNet uses fixed-width convolutional layers, two downsampling stages, minimal skip connections, and nearest-neighbor upsampling.

On the training side, TroyNet uses a custom loss function that gives much more weight to P- and S-phase picks, helping the model learn from rare but critical labels and focus on the most important parts of the signal. Training used a cosine-annealing learning-rate schedule with warm restarts (SGDR; Loshchilov & Hutter, 2017). This approach works well for problems like seismic phase picking, where P and S labels are extremely rare compared with background noise. The periodic restarts allow the optimizer to jump out of poor local minima and keep learning steadily even as the loss landscape becomes complicated (Buda et al., 2018; He et al., 2019). Because our models train for a relatively long time (150 epochs), SGDR helped maintain stable progress throughout the later stages of training.

These combined strategies allow TroyNet to deliver accurate and reliable phase picks in noisy, real-world conditions. TroyNet's performance is compared to retrained PhaseNet baseline implemented and evaluated under identical conditions and to hand-picked phase catalogs and also on continuous data. Finally parameters for the earthquakes resolved using the P and S wave arrival duration database were used to create the a 1-D crustal model to compare performances of algorithms against expert picks. In summary, this study contributes: (i) a focused design while introducing a single transformer bottleneck, combined with parameter-free sinusoidal positional encoding to capture long-range temporal context, (ii) a custom loss function emphasizing rare but critical classes, and (iii) extensive evaluation on a large regional dataset from northwestern Türkiye.

2. Data & Methods

2.1 Data

The TroyNet model was tested on a seismic dataset gathered from northwestern Türkiye. The area is covered by a dense seismic network of stations run by Çanakkale Onsekiz Mart University (COMU) and national institutions, such as the Türkiye Disaster and Emergency Management Presidency (AFAD) and Kandilli Observatory and Earthquake Research Institute (KOERI). Our dataset consists of a waveform database of local earthquakes between 2008 and 2022 from the KOERI catalog with manually detected pick times and covers the Biga Peninsula and nearby regions (Figure 2). Manual P- and S-phase picks have been produced by a team of professional analysts from KOERI. Individual analyst identifiers are not available, and inter-analyst variability cannot be quantified directly. The database contains 62,808 earthquakes with local magnitudes ranging from approximately 0.4 to a maximum of 6.3. These events were registered by 242 stations. The dataset contains 190,235 waveforms with both P and S arrivals. Each waveform consists of three-component recordings (E, N, Z) around the time of the earthquake. For consistency, all of the waveform traces were resampled at 100 Hz and bandpass filtered (1–45 Hz). Baseline correction and normalization were performed on every trace via mean subtraction and division by standard deviation, standardizing the amplitude scales across the dataset. The original KOERI seismic recordings are approximately 5 minutes long. Windows of 30 seconds, starting a random few seconds prior to the hand-labeled P-wave arrival, were employed for training and testing.

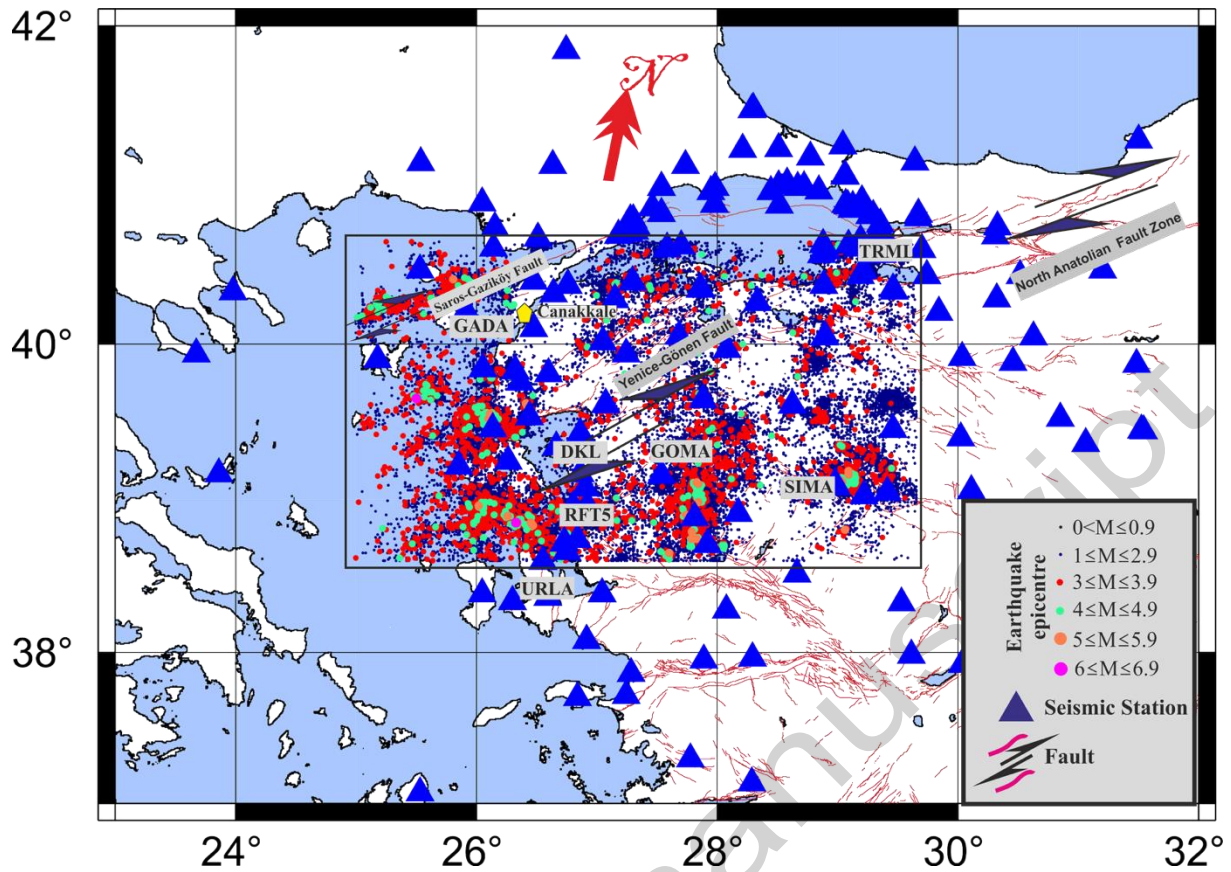


Figure 2. Spatial distribution of earthquakes and seismic stations used in this study across northwestern Türkiye. Earthquakes are shown according to their magnitude classes, and seismic stations that recorded these events are indicated. Stations included in the analysis are highlighted. Mapped active faults and their names are also shown for tectonic reference. The study region is shown as black rectangle.

To analyze data, the RMS-based SNR distribution is computed on a logarithmic decibel (dB) scale from pre- and post-P-arrival amplitudes (Figure 3). It shows a strong positive skew. Most traces cluster between 0 dB and 20 dB, and there is a modest tail above 40 dB. Although truly high-SNR records are only 0.67% above 50 dB, they can still bias training by encouraging the model to over-fit to these exceptionally clean signals. At the opposite end, 3.39% of traces fall below 0 dB; these very noisy examples add little useful information yet introduce instability in early training. To moderate both extremes, we clip SNR values for training (and validation) data to the 0–50 dB range, sacrificing only 4.06% of the data in total while stabilizing the distribution at 12.85 ± 10.19 dB (Figure 2). The test data were not SNR clipped for testing under real-life conditions.

To discourage the model from learning a constant temporal shift for P arrival, a random pre-arrival shift of 2–6 s was introduced. This period could have been extended, but it was chosen from the observed variation in pre-event signal characteristics, with more than 48% of events in our catalog having SNR (dB) < 10, ensuring sufficient noise variety to prevent overfitting to deterministic arrival locations (Figure 3). By incorporating random offsets, the model is forced to rely on signal features instead of positional bias, thereby increasing robustness and performance in the field.

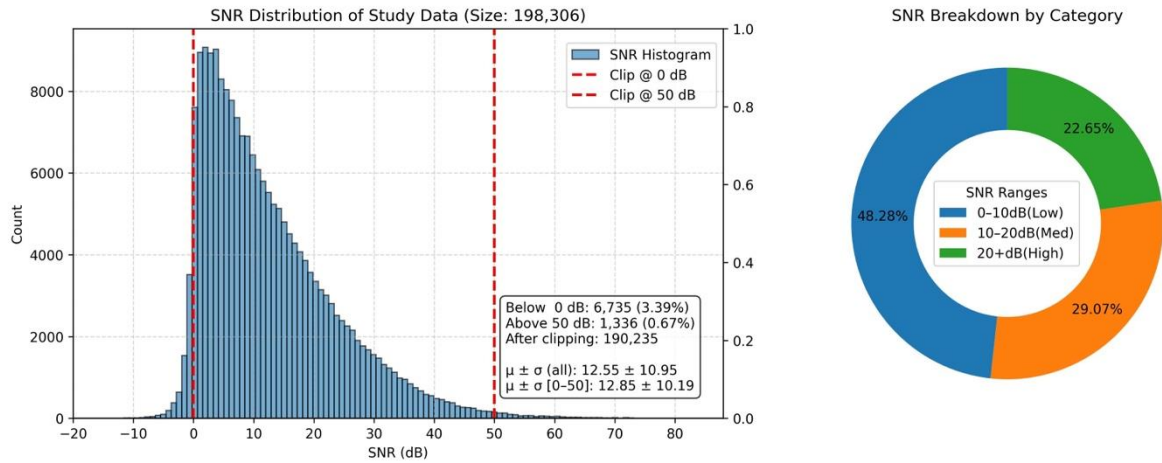


Figure 3. RMS-SNR (dB) distribution. On the left, RMS-SNR (dB) histogram of the 190,235 labelled three-component seismograms. Red dashed lines mark the 0 dB and 50 dB clipping thresholds for training data; 3.39% of seismograms fall below and 0.67% above this range. On the right, pie chart of the same dataset showing the share of seismograms in each SNR tier (low, medium, high) before clipping.

For every earthquake, the P- and S-wave arrival times were selected by skilled analysts and used as the training and testing ground-truth labels. This dataset was divided into the following sets: $\sim 70\%$ (133K) of the events for training, $\sim 15.3\%$ (29K) for validation (for tuning model parameters), and $\sim 14.7\%$ (28K) for testing final performance.

2.2 Architecture

TroyNet is a U-Net-style encoder-decoder network for time series sequence labeling of seismic data and is specifically tailored for seismic phase picking: the input seismogram is downsampled with fixed-width convolutional layers, temporal patterns are captured, and a symmetrical upsampling path produces a per-time-step probability for the P phase, S phase, and noise. The encoder begins with two [Conv1D(64, 3) $\rightarrow$ BatchNorm $\rightarrow$ ReLU] blocks. Six dilated residual units span multiple temporal scales. Before entering the transformer, a parameter-free sinusoidal positional encoding is added. This helps the subsequent 8-head self-attention layer, which has LayerNorm and 20% dropout, learn long-range dependencies that plain convolutions might miss. The original temporal resolution is restored by two nearest-neighbor UpSampling1D stages, and each stage is followed by Conv1D(64, 3).

TroyNet uses just two skip connections instead of a complete U-Net cascade. The first skip, drawn right after the first pooling step, reintroduces high-resolution features that are crucial for sharpening arrival picks close to the input scale (fine temporal context). Since this skip is drawn from a downsampled representation, it is upsampled prior to concat operation to achieve dimensional consistency with the decoder. It is an intentional compromise that maintains efficiency while still injecting fine-scale information. The second skip, drawn after the dilated residual stack, injects again mid-scale features that capture multi-temporal patterns over several seconds. Together, these two carefully placed skips inject complementary detail without the complexity of multiple cropping, alignment, and concatenation operations characteristic of deeper U-Net architectures. The final 1×1 convolution outputs a three-class softmax at each point of the 3000 samples, which corresponds to a 30 s window at 100 Hz (Figure 4).

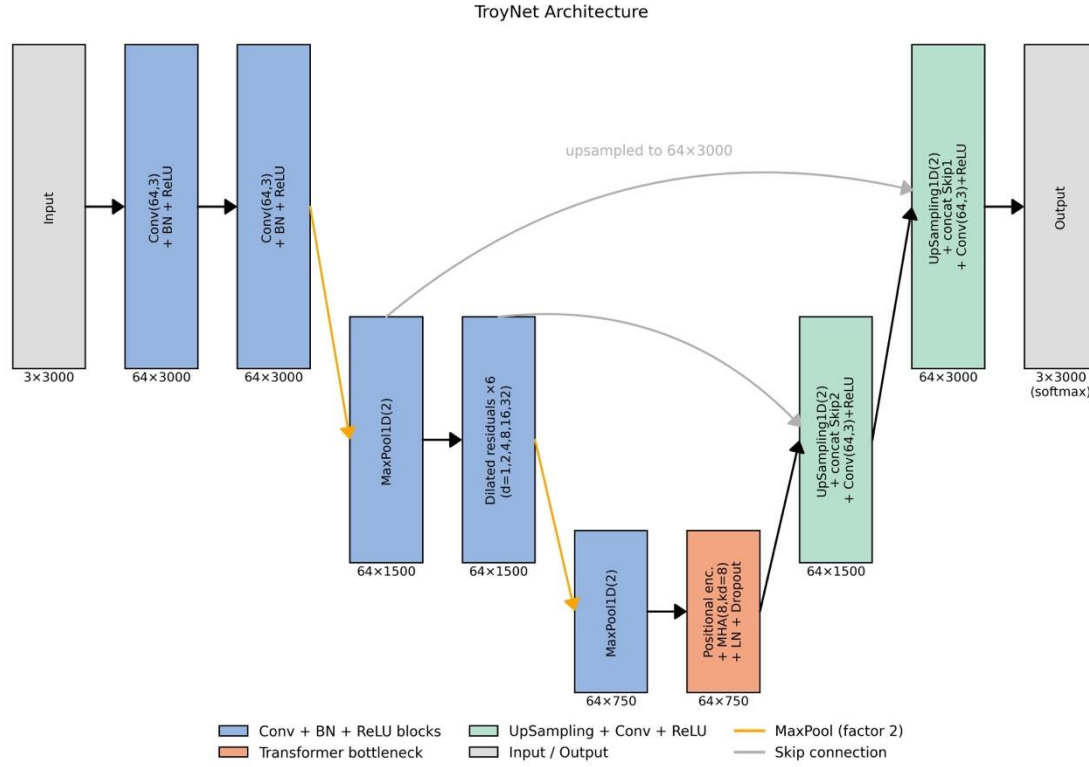


Figure 4: Boxes have dimensions in the form of "channels $\times$ samples." Input has three components (E, N, Z) across 3000 samples. Output is three probability classes (Noise, P, S) per point across the same 3000 samples. The encoder begins with two convolutional blocks. A pooling layer and a stack of six dilated residual blocks follows to model multi-scale temporal patterns. A second pooling stage compacts the sequence to 750 steps. Then a lightweight transformer bottleneck with sinusoidal positional encoding and multi-head self-attention introduces long-range temporal context. The decoder restores temporal resolution through two upsampling stages, each followed by convolutional refinement and skip concatenation. To retain efficiency, there are only two skip connections: one from the first pooling stage, which is upsampled before concatenation to restore fine-scale temporal detail, and one from the residual stack, which reintroduces mid-scale features. Together, these skips add complementary fine and mid-level context without the complexity of a full U-Net cascade.

In the TroyNet architecture, after the convolutional and residual blocks, features are input into a transformer block. Sinusoidal positional encodings (Vaswani et al., 2017) are added to the feature sequence before the transformer block to encode absolute timing information and enable the self-attention mechanism to infer relative time lags. Since the transformer itself does not inherently know the position of each timestep (it only sees feature relationships), to ensure that the model retains knowledge of when something is happening in the sequence, positional information is injected by adding a parameter-free sinusoidal positional encoding (PE) to the sequence. The input shape after the second MaxPooling becomes (750, 64) before the transformer. Therefore, for an integer time index, $\text{pos} \in [0, 749]$, and integer embedding dimension " i " $\in [0, D-1]$ (where D is the channel width and equals 64), encoding is explained in equation (1).

$$\text{PE}_{2i} = \sin(\text{pos}/10000^{2i/D}), \text{PE}_{2i+1} = \cos(\text{pos}/10000^{2i/D}) \quad (1)$$

These sinusoids of different frequencies allow the model to recognize absolute positions and understand the relative distances between picks. Adding (rather than concatenating) PE to the feature map introduces no extra trainable parameters. As a result, the shift-equivariance established by the dilated convolutions is preserved, so the self-attention layer causes both absolute and relative time lags.

The final TroyNet configuration contained 231,043 parameters and 229,251 of them were trainable. These numbers corresponded to a weight memory footprint of ~ 0.9 MB in FP32. All experiments were conducted in Google Colab using an NVIDIA T4 GPU. During training (batch size 32; 3000×3 inputs), the step time was ~ 0.13 – 0.14 s per step (~ 9 – 10 min per epoch for 4159 steps). Inference speed was evaluated on the same GPU over the full test set (28,008 windows), yielding ~ 1.07 ms per window (~ 934 windows/s) for TroyNet, compared to ~ 0.27 ms per window ($\sim 3,736$ windows/s) for the retrained PhaseNet baseline.

To combat severe class imbalance in seismic phase picking—where an overwhelming majority of the time data samples belong to the background class—a sparse categorical cross-entropy loss function that manually sets class weights was used. The loss function places far greater significance on P- and S-phase picks (classes 1 and 2), both of which are weighted much more (1 to 2998) over the background class (class 0). This allows our model to be biased into making correct phase pick predictions despite these not being very common in the training data. This strategy is analogous to class-weighted loss functions commonly used in rare-event modeling (King & Zeng, 2001). This is complementary to our sample-level SNR-based weighting and discourages the model from collapsing to always predict the majority class.

The PhaseNet baseline used here was re-implemented in TensorFlow 2 / tf.keras in terms of true 1-D convolutional operations, rather than the pseudo-1D (2-D convolutions with $[k, 1]$ kernel and stride sizes) in the original TensorFlow 1.x release. This update keeps the architecture exactly: U-Net encoder–decoder depth, filter progression, kernel sizes, downsampling and upsampling strategy (strided convolution and transposed convolution), skip connections, center-crop concatenation for alignment, and final 1×1 convolution with softmax activation are identical to the original. The 1-D version is mathematically equivalent to the original pseudo-1D approach and has the same output shapes and parameter counts for corresponding layers, allowing an even and direct comparison. For a fair comparison, PhaseNet was retrained using the same preprocessing steps, data splits, optimizer, learning-rate schedule, loss weighting, and evaluation criteria as the proposed model. This ensures that any performance differences are not driven by training or evaluation choices.

2.3 Methods

2.3.1 Training

The training set was made up of actual earthquake recordings only from the regional network. While no independent noise-only windows were included, most samples within these event windows have low SNR, so the model is exposed naturally to a large range of non-event conditions without the addition of synthetic noise segments.

To understand the role of each architectural component in TroyNet, a step-by-step ablation study has been planned. The full model was gradually complexified, with each extended version adding only one element at a time while keeping all other training settings identical.

In total, four variants of TroyNet have been evaluated along PhaseNet: a base model containing only the dilated-residual encoder-decoder structure (TroyNet-Base), the base model with the transformer bottleneck added (TroyNet-Base+Transformer), the next model to test also adds positional encoding in the transformer block (TroyNet-Base+Transformer+PE) and finally the last model utilizes SNR-based training on top of all architectural improvements till this point (Full TroyNet).

To address different levels of noise in the training data, the SNRs (dB) for all waveforms based on the RMS amplitude prior to and following the P-wave arrivals were calculated. Very low and high values of the SNR (dB) were clipped to avoid the effects of very noisy or too clean samples (Figure 3). For the Full TroyNet model in ablation study, the best weighted discretization strategy was then applied after testing different weight combinations: Samples with SNRs <5 were weighted 1.0, samples with SNRs within $[5, 10]$ were weighted 1.7, samples with SNRs within $[10, 20]$ were weighted 2.3, and samples with SNRs >20 were weighted 3.0. This strategy prioritizes cleaner examples to make the model learn picking patterns but also incorporates large number of noisy events with a lesser effect.

Optimization was performed via SGDR (Loshchilov & Hutter, 2017), which decreases the learning rate periodically to enable smoother convergence and aids in avoiding local minima. TroyNet was trained for 150 epochs with a cycle length of 15 epochs. To improve the pick localization, parabolic interpolation to softmax probability peaks was applied during the evaluation phase to increase the temporal precision of TroyNet's predicted arrivals.

The target labels are discrete, with 1 marking the P-pick sample and 2 marking the S-pick sample, and 0 elsewhere (class 0: noise). No label smoothing or Gaussian/triangular target shaping was applied during training. Gaussian kernels with varying σ (along with other parameters) were only used during evaluation to optionally smooth the model's predicted probability sequences before peak detection as explained below.

2.3.2 Determination of Positive Pick Threshold

To arrive at a confidence criterion (positive pick thresholds) that does not depend on a specific model, we first looked at how confident the networks actually are in practice using validation dataset. We examined the empirical distributions of the maximum softmax probabilities produced by a TroyNet model (We selected representative TroyNet Base+Transformer architecture) and PhaseNet by treating P and S phases separately. Figure 5 shows the cumulative distributions of these maximum values. The shape of the curves illustrates how the probability mass gradually accumulates. For both models, P-phase probabilities tend to rise sharply beyond about 0.95, while S-phase probabilities increase more slowly and peak at slightly lower values. Since S arrivals are usually less impulsive and are more affected by noise and waveform variability, this behavior is expected. PhaseNet exhibits an earlier accumulation of high P-phase probabilities, whereas TroyNet shows a sharper transition at higher confidence levels, indicating a more conservative probability calibration. A common P-phase threshold of 0.97 was selected to represent a shared high-confidence regime, avoiding model-specific optimization and enabling a consistent comparison between the two approaches. For the S phase, the cumulative distributions of TroyNet and PhaseNet show similar behavior with a more gradual transition compared to the P phase. A slightly lower threshold (0.95) was therefore adopted to account for the broader and lower-SNR nature of S arrivals, balancing timing accuracy and pick availability without favoring either model. These values were not tuned to maximize any specific metric, but were chosen to

define a consistent high-confidence regime. The same thresholds were applied to both TroyNet and PhaseNet without modification. All timing metrics (MAE and RMSE) and detection metrics (precision, recall, and F1) were computed after applying this positive pick thresholding steps.

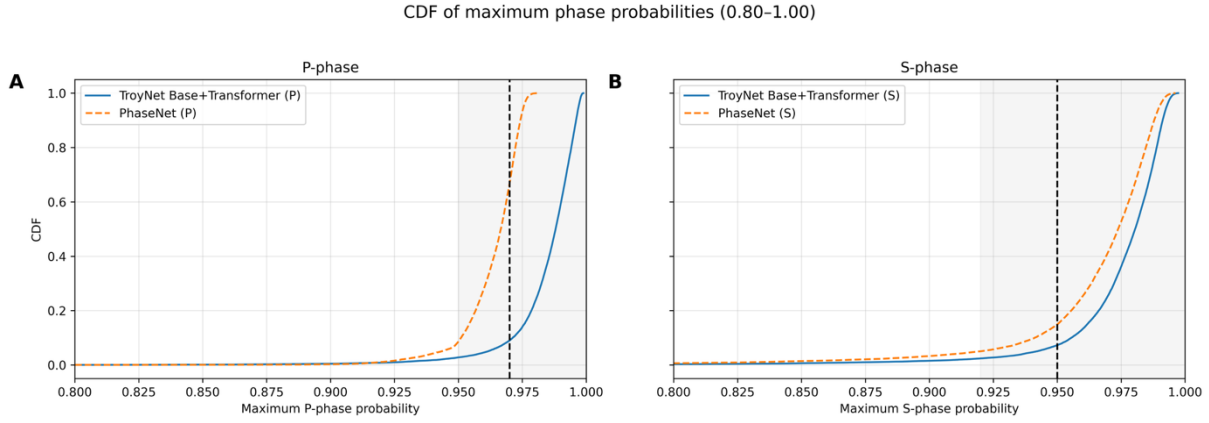


Figure 5: Cumulative distributions of the maximum softmax probabilities (CDFs) for P- and S-phase picks produced by TroyNet and PhaseNet on the validation dataset. The shaded regions indicate the high-confidence probability ranges ([0.95,1.0] for P-phases and [0.92,1.0] for S-phases). Vertical dashed lines mark the selected phase-specific positive pick thresholds (0.97 for P, 0.95 for S), which are applied uniformly to both models. The CDFs show how probability mass concentrates rapidly near unity. This fact implies the use of fixed high-confidence thresholds. Similar CDF values at the selected thresholds do not imply comparable picking performance, as timing accuracy and false detections are evaluated separately.

As a result, confusion matrix as shown in Table 1 has been build, model picks whose softmax score reached at least 0.97 for P-picks / 0.95 for S-picks was called a positive detection—and to be counted as a true positive it also had to fall within ± 0.2 seconds (± 20 samples, 100Hz) of the manual pick. Although smaller windows such as ± 10 samples are commonly reported, a wider tolerance was considered more appropriate for local and noisy recordings, where phase onsets, especially for S waves, are often less impulsive and more spread in time. The network also produces smooth probability peaks rather than single-sample impulses, and small shifts around the peak are therefore expected after smoothing and interpolation. Using ± 20 samples provides a more realistic acceptance range without affecting the relative ranking of the models..

Confusion Matrix		Predictions	
		Positive Predictions TroyNet P Pick Prob. ≥ 0.97 TroyNet S Pick Prob. ≥ 0.95	Negative Predictions TroyNet P Pick Prob. < 0.97 TroyNet S Pick Prob. < 0.95
Actual Labels	Actual Positive	TP (True Positive) $ \Delta t \leq 0.2$ s	FN (False Negative) $ \Delta t \leq 0.2$ s
	Actual Negative	FP (False Positive) $ \Delta t > 0.2$ s	TN (True Negative) $ \Delta t > 0.2$ s

Table 1: Confusion matrix used for classification in this study. A time window of ± 0.2 s was selected, as this accuracy is acceptable for real-world seismogram analysis.

2.3.3 Need for Gaussian Smoothing

The model outputs a probability curve for the P and S classes for each waveform as shown in Figure 6. The probability curves produced by the network are often smooth, broad, or slightly asymmetric rather than spike-like. This behavior arises in part because the loss function encourages the model to distribute probability around the most likely onset when the exact sample is uncertain. It is also a consequence of the training labels, which mark only a single sample as “1” (P) or “2” (S) and assign zero elsewhere, forcing the network to infer a gradual transition rather than a perfectly sharp boundary. As a result, the raw softmax output typically forms a peaked region rather than a single-sample impulse. To solve the problem, a light Gaussian smoothing was applied to suppress isolated spikes without altering the general shape of the peak. The pick itself was then taken from the dominant maximum, and its timing was refined by fitting a small parabola to the peak and its two immediate neighbours. This step provided a sub-sample estimate of the arrival time and offered a more stable representation of the peak position than relying on the raw discrete index alone.

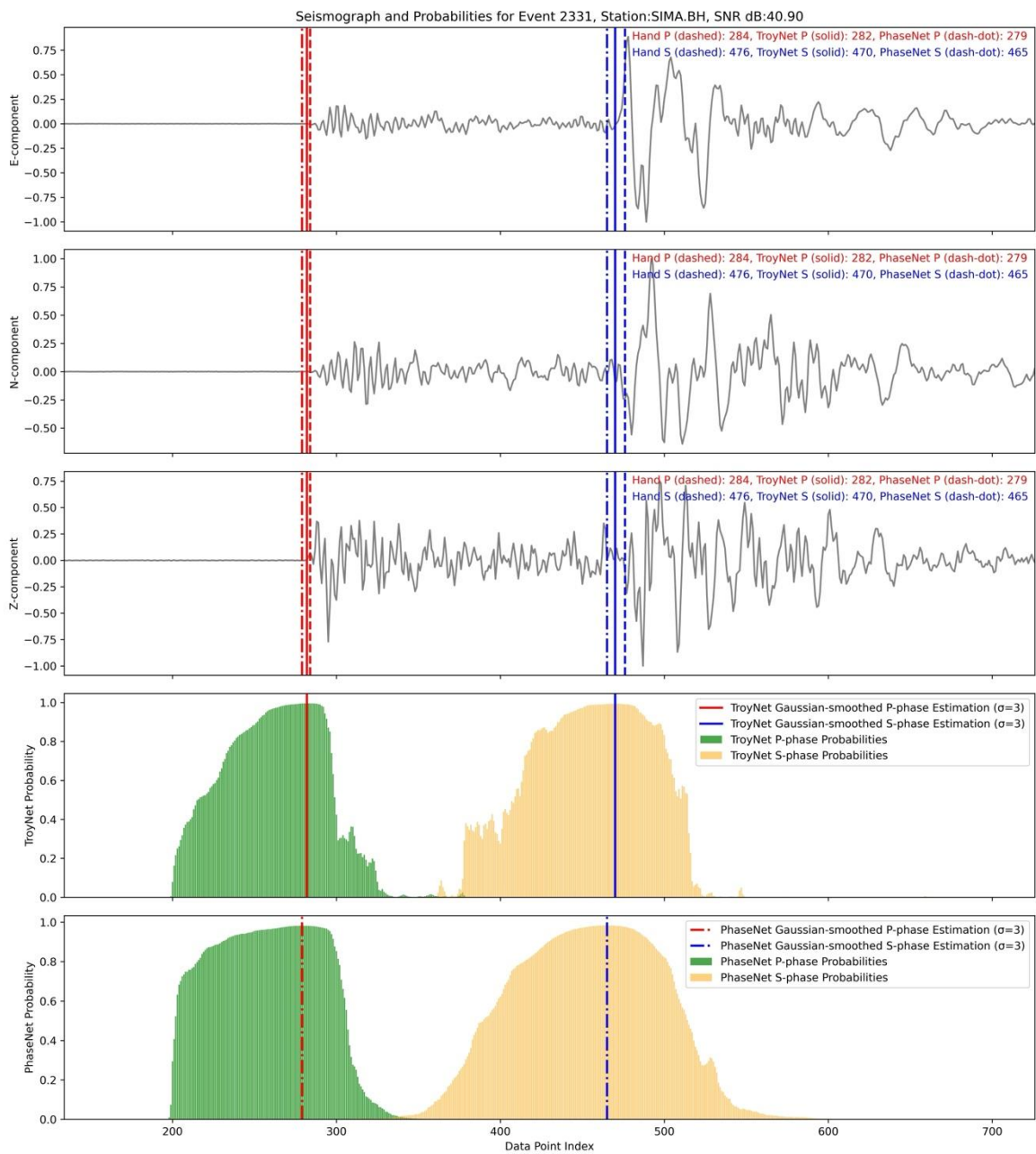


Figure 6: Example high-SNR seismogram and TroyNet phase pick predictions for a local earthquake ($M = 3.0$) at station SIMA.BH on June 9, 2012. Epicenter of the event is 39.16° N, 29.09° E (Simav–Söğüt, Kütahya), 12.6 km away from the SIMA station. The first three panels display the E, N, and Z components of ground motion. The fourth panel shows TroyNet's predicted softmax probability distributions for the P-phase (green) and S-phase (blue) phases smoothed in time with a Gaussian filter ($\sigma = 3$) and re-normalized, matching the evaluation configuration. The manual picks are dashed lines (red for P, orange for S), and TroyNet's predicted picks are solid lines. The time is the number of data points (1 point = 0.01 s) from the start of the recording. Both phases are correctly identified by the model, showing its strength under low-noise conditions.

After positive pick thresholds have been defined, varying levels of Gaussian smoothing ($\sigma = -1$ and 1 to 5, -1 meaning no smoothing) have been examined. Beyond $\sigma = 3$, additional smoothing provided diminishing returns. 0.95 for S-picks and 0.97 for P-picks together with Gaussian smoothing $\sigma = 3$ allows a good compromise between RMSE and F1 and provides a satisfactory number of predictions. These high prediction rates are particularly important in real-life applications such as automated earthquake early warning systems and real-time seismic monitoring, where both accurate detection and precise timing are critical (Figure 7).

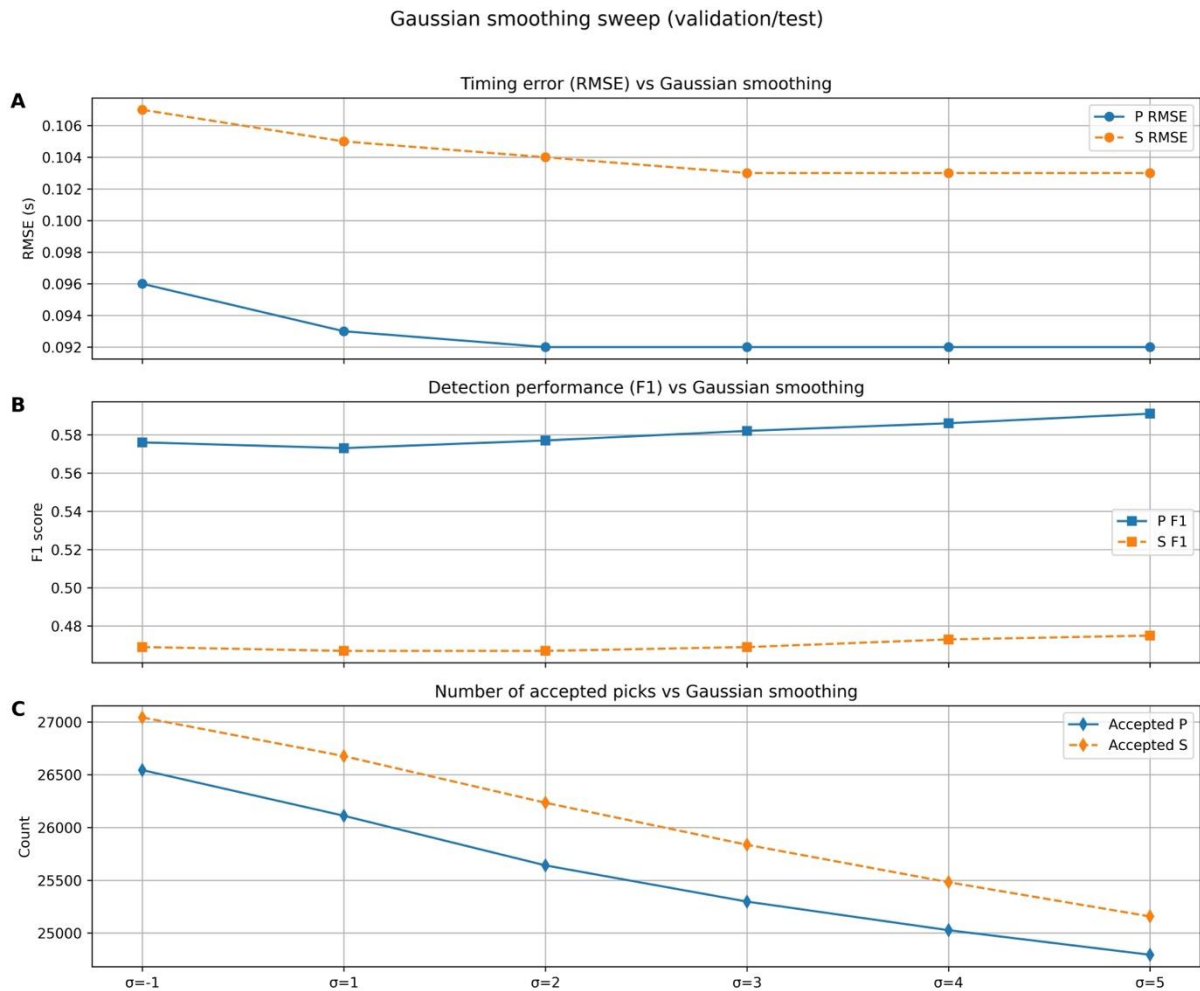


Figure 7. Effect of Gaussian smoothing on TroyNet phase-picking performance evaluated on the validation dataset using fixed confidence thresholds ($P \geq 0.97$, $S \geq 0.95$). The three panels show (A) the average RMSE of P- and S-phase picks, (B) the corresponding F1 scores, and

(C) the number of accepted P and S picks. Increasing the smoothing level improves timing stability and detection consistency up to $\sigma \approx 3$, beyond which further smoothing yields negligible gains while reducing the number of accepted picks. The vertical dashed line marks the smoothing level ($\sigma = 3$) adopted for all subsequent evaluations. Picks with softmax probabilities below the fixed thresholds are treated as no-pick cases.

2.3.4 Ablation Study

The evaluation protocol (confidence thresholds and Gaussian smoothing) was defined once using the validation dataset and fixed before the ablation study. No further parameter tuning was performed, and the validation data were used only for consistent model and epoch selection. To understand the role of each architectural component in TroyNet, a step-by-step ablation study has been carried-out. The full model was gradually complexified, with each extended version adding only one element at a time while keeping all other training settings identical. In total, three variants have been evaluated: a base model containing only the dilated-residual encoder-decoder structure (TroyNet-Base), the base model with the transformer bottleneck added (TroyNet-Base+Transformer), the next model to test also adds positional encoding in the transformer block (TroyNet-Base+Transformer+PE) and finally the last model model utilizes SNR based training on top of all architectural improvements till this point (Full TroyNet). This design allowed for the examination of the individual contribution of each additional block independently. All models were trained with the same data split, SGDR schedule, optimizer, and class-weighted loss. During training, all models were logged after each epoch for subsequent evaluation. Following standard practice in seismic phase-picking evaluation, we selected the best epoch based on the minimum Mean RMSE ($[\text{RMSE}_P + \text{RMSE}_S]/2$) among the top-performing models ranked by validation RMSE. When two candidates had Mean RMSE, we selected the one with higher Mean F1 ($[\text{F1}_P + \text{F1}_S]/2$) as a secondary tie-breaker. The numerical results of these controlled comparisons are presented in the Results section.

After the ablation study, selected final TroyNet model and baseline implemented and evaluated under identical conditions as explained in introduction has been compared using test data and a small continuous data. Finally parameters for the earthquakes resolved using the P and S wave arrival duration database were used to create a 1-D crustal model to compare performances of algorithms against expert picks.

3. Results

3.1 Ablation Study Results

To quantify the incremental benefit of each architectural component, several progressively enhanced TroyNet variants on the full validation set using a fixed evaluation protocol ($P \geq 0.97$, $S \geq 0.95$, Gaussian smoothing $\sigma = 3$, tolerance ± 0.20 s) have been evaluated. Under this TP-based timing definition, all variants show similar mean timing performance (Mean RMSE ≈ 0.095 s), which suggests that the core dilated-residual encoder-decoder already captures most of the onset timing information. The differences become more visible in pick quality and stability metrics (F1, precision/recall, and the number of accepted picks). Adding the transformer bottleneck led to a small but consistent improvement in S-phase behavior: S-RMSE decreased ($0.1025 \rightarrow 0.1012$ s) and S-F1 increased ($0.4826 \rightarrow 0.4964$), together with a higher number of accepted picks. Adding positional encoding yielded the highest number of accepted S picks (up to 27,345), indicating a more permissive but stable pick selection.

However, this did not translate into a clear improvement in S timing metrics (S-RMSE $\approx$ 0.103 s). Overall, the best balance between timing and detection across phases is obtained by the transformer-based variants. The effect of positional encoding is reflected more strongly in coverage (accepted picks) than in timing. SNR-aware training produced only slightly shifted precision–recall trade-offs relative to the other variants. A clear reduction in the TP-based timing errors has not been observed. This suggests that the main gains in our setting are driven by architectural context modeling. As a result TroyNet Base + Transformer + PE architecture has been selected as final TroyNet. A broader comparison of alternative SNR weighting strategies is left for future work.

Metric	TroyNet Base	TroyNet Base + Transformer	TroyNet Base + Transformer + PE	Full TroyNet
Mean RMSE (s)	0.0955	0.0951	0.0955	0.0953
P MAE (s)	0.0713	0.0716	0.0708	0.0712
S MAE (s)	0.0859	0.0848	0.0866	0.0848
P RMSE (s)	0.0886	0.0890	0.0880	0.0891
S RMSE (s)	0.1025	0.1012	0.1031	0.1016
P Precision	0.6910	0.6728	0.6738	0.6746
P Recall	0.5226	0.5390	0.5546	0.5681
P F1	0.5951	0.5985	0.6084	0.6168
S Precision	0.5339	0.5333	0.5099	0.5249
S Recall	0.4403	0.4643	0.4781	0.4719
S F1	0.4826	0.4964	0.4935	0.4970
P Bias (s)	0.0116	0.0149	0.0123	0.0137
S Bias (s)	0.0043	0.0051	0.0009	0.0040
Accepted P picks	22062	23368	24006	24560
Accepted S picks	24054	25394	27345	26220

Table 2: Ablation results for TroyNet variants on the validation set using the fixed evaluation protocol (Gaussian smoothing $\sigma = 3$; thresholds $P \geq 0.97$ and $S \geq 0.95$; tolerance ± 0.20 s). Reported timing metrics (MAE, RMSE, bias) are computed on true-positive picks only, while “Accepted picks” denotes the number of picks exceeding the confidence threshold.

3.2 Final TroyNet vs PhaseNet Comparison

Selected TroyNet version after ablation study was compared with PhaseNet which was retrained with the same training dataset. We used a PhaseNet-style U-Net architecture retrained on the same dataset with the same loss function and SGDR schedule used for TroyNet, enabling a controlled architecture-only comparison. For this comparison, the full test dataset has been used. The comparison was performed both on the complete dataset (0–50 dB SNR) and on three SNR-defined subsets (<10 dB, 10–20 dB and ≥ 20 dB). Table 3 summarizes the results.

Across the full test set, the two models show broadly comparable timing performance across all SNR ranges. The clearest separation is in pick completeness for both phases. TroyNet accepts more high-confidence picks: on the full test set, it retains 81% of P picks at probabilities ≥ 0.97 , whereas PhaseNet retains only 69%. This gap is even larger for S picks (94% vs 71% for S-pick probabilities ≥ 0.95). This difference is most pronounced in the low-SNR band, where TroyNet still recovers 90% of S picks and reaches an S-phase recall of 0.37, while PhaseNet drops to 58% recovery with an S-phase recall of 0.29. In the high-SNR bin, both models perform strongly, but TroyNet still produces more usable high-confidence picks.

Overall, TroyNet achieves slightly lower P-phase timing errors, while PhaseNet shows marginally lower S-phase timing errors across all SNR bins.

Metric	Final TroyNet				Retrained PhaseNet			
	0-50 dB	0-10 dB	10-20 dB	20+ dB	0-50 dB	0-10 dB	10-20 dB	20+ dB
P MAE (s)	0.0714	0.0840	0.0738	0.0589	0.0744	0.0857	0.0789	0.0634
S MAE (s)	0.0841	0.0894	0.0831	0.0787	0.0835	0.0882	0.0837	0.0788
P RMSE (s)	0.0894	0.1012	0.0913	0.0764	0.0907	0.1022	0.0952	0.0783
S RMSE (s)	0.1009	0.1056	0.1004	0.0955	0.1002	0.1049	0.1003	0.0953
P Precision	0.6642	0.5018	0.7002	0.8436	0.7239	0.5648	0.7333	0.8576
S Precision	0.5140	0.4101	0.5559	0.6647	0.6023	0.5088	0.6232	0.7077
P Recall	0.5363	0.3191	0.6742	0.8320	0.4979	0.2390	0.6675	0.8445
S Recall	0.4808	0.3674	0.5358	0.6551	0.4297	0.2940	0.5009	0.6322
P F1	0.5934	0.3901	0.6869	0.8378	0.5900	0.3358	0.6988	0.8510
S F1	0.4968	0.3876	0.5457	0.6599	0.5016	0.3727	0.5554	0.6678
P Bias	0.0178	0.0149	0.0154	0.0226	0.0009	0.0050	-0.0026	0.0017
S Bias	0.0045	0.0076	0.0051	0.0004	0.0008	0.0046	-0.0003	-0.0020
P Picks (Prob ≥ 0.97)	80.74%	63.60%	96.28%	98.63%	68.78%	42.31%	91.03%	98.47%
S Picks (Prob ≥ 0.95)	93.54%	89.59%	96.37%	98.55%	71.34%	57.78%	80.38%	89.34%

Table 3: Comparison of TroyNet (final architecture) and retrained PhaseNet on the test dataset, shown for the full dataset and for three SNR ranges (<10 dB, 10–20 dB and ≥ 20 dB). Plotted metrics include P- and S-phase MAE, RMSE, precision, recall, F1 score, timing bias and the percentage of high-confidence picks. TroyNet shows similar MAE/RMSE and achieves higher recall and better S-phase pick completeness across all SNR bins.

Overall, these results suggest that the most substantial advantage of TroyNet is its robustness to noise and its ability to recover phases reliably across the full range of SNR conditions. While PhaseNet maintains slightly lower timing errors overall, TroyNet produces far more complete pick sets and substantially higher phase recall. Northwestern Türkiye regional network has high noise levels, and S-phase availability often limits event location accuracy. The improved reliability and consistency of TroyNet are likely to yield more complete catalogs with stable earthquake locations.

The slightly wider confidence intervals for PhaseNet, particularly for S-phase RMSE, primarily reflect the smaller number of true-positive S picks compared to TroyNet, which leads to increased variability in the bootstrap resamples. Confidence intervals for MAE and RMSE were estimated using bootstrap resampling (1000 realizations) of true-positive pick timing errors, and 95% intervals were derived from the 2.5th and 97.5th percentiles.

Metric	Final TroyNet	Retrained PhaseNet
	0-50 dB	0-50 dB
P MAE (s)	0.0714 [0.0705–0.0722]	0.0744 [0.0735–0.0753]
S MAE (s)	0.0841 [0.0832–0.0850]	0.0835 [0.0825–0.0845]
P RMSE (s)	0.0894 [0.0885–0.0902]	0.0907 [0.0898–0.0916]
S RMSE (s)	0.1009 [0.0999–0.1017]	0.1002 [0.0993–0.1011]

Table 4: Uncertainty of timing performance for TroyNet and retrained PhaseNet on the test dataset. Values are reported with 95% bootstrap confidence intervals computed from true

positive pick timing errors using the same smoothing, tolerance, and phase-specific pick thresholds ($P \geq 0.97$, $S \geq 0.95$) as in Table 3, based on 1000 bootstrap resamples.

To also validate the temporal coherence of the phase picks, the distributions of the absolute errors for the P and S phases are plotted in Figure 8. On the left, although the P-phase error distribution was less widespread (lower variance), the average signed timing error (bias) of the P-phase was slightly greater than that of the S-phase. This means that P picks are consistently offset from the true pick location but with less variability. The S-phase picks had lower average bias but greater variability, which is likely due to the emergent and noisy nature of S arrivals. This is in accordance with seismological expectations: P onsets are impulsive but may be systematically delayed (e.g., from smoother peaks after Gaussian filtering), whereas S-phase picks are more dispersed but sometimes better centered. On the right, each SNR-segmented histogram is shown as a line plot to prevent the visual clutter of three overlaid bar charts: Blue corresponds to high-SNR traces (> 20 dB), green to moderate SNR (10–20 dB), and red to low SNR (< 10 dB). Solid lines show P-phase errors and matching dashed lines show S-phase errors, so readers can compare how timing bias (small rightward shift) and spread (broader tails) evolve from high to low SNR.

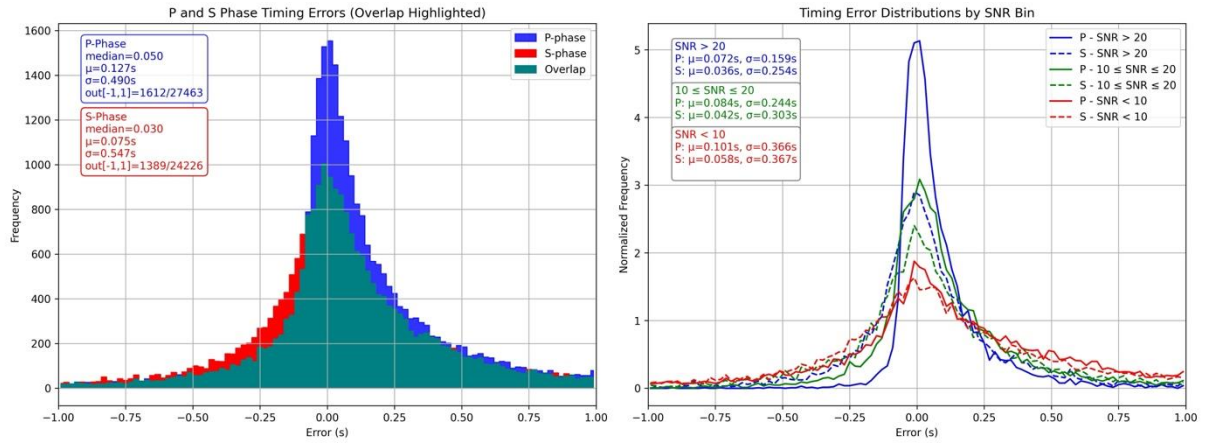


Figure 8: Signed timing errors of TroyNet picks. On the left, overlaid histograms of P-phase (blue) and S-phase (red) residuals on the full NWTR test set. Median, mean (μ), and standard deviation (σ) are annotated and counts outside ± 1 s are listed. On the right histograms of timing error for three RMS-SNR (dB) classes: low (< 10 dB, blue), moderate (10–20 dB, green) and high (≥ 20 dB, red). Legends report mean and σ for each subset.

3.3 Performance on Continuous Dataset

TroyNet is designed as a phase picker rather than an event detector; therefore, its continuous-data performance is evaluated only within manually identified earthquake intervals. Developing a dedicated detection module to operate jointly with TroyNet is beyond the scope of this study, but such detector–picker separation is standard in operational seismology, where triggering and phase picking are handled by distinct algorithms.

To examine how TroyNet behaves on realistic continuous waveforms, an additional test using two 2-hour, three-component recordings from the KO network was performed: station GADA (08 January 2022, 12:00–14:00 UTC) and station DKL (06 January 2022, 19:00–21:00 UTC). Each 2-hour segment contains two local earthquakes with manually reviewed P- and S-arrivals (four events, eight picks in total) (Figure 9). The continuous data were preprocessed

identically to the training set (1–45 Hz band-pass and 100 Hz sampling) and divided into 30-s windows with a 10-s step (66% overlap). There were total of 1436 windows: 12 windows with P handpicks, 12 windows with S handpicks, 16 windows with any label and 1420 noise windows.

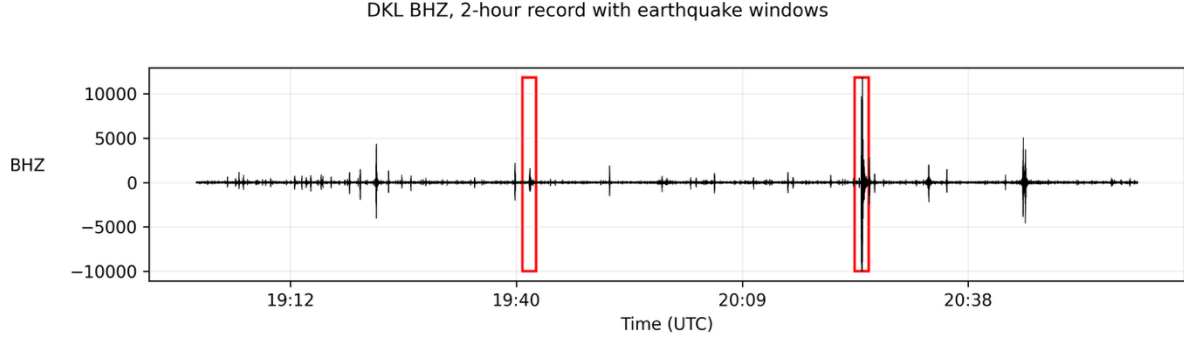


Figure 9: Two-hour continuous vertical-component recordings from station DKL, with the time intervals containing the two local earthquakes highlighted in red. These segments were used to evaluate TroyNet under realistic continuous-data conditions.

Our first test scenario for continuous tests consists of isolated, event-centered windows. Because continuous data contain many overlapping windows and the true pick may lie in several of them, we used a window-aggregation strategy: TroyNet’s pick time in each window using parabolic interpolation around the peak softmax probability. The error for that event was defined as the smallest absolute difference among all valid windows. This scheme isolates the timing accuracy of the picker itself and avoids counting multiple overlapping predictions as separate picks. Across the eight phase arrivals, TroyNet achieved consistent timing performance, with absolute errors ranging from a few milliseconds to $\sim 2\text{--}3$ s depending on local SNR (overall MAE = 1.171 s, RMSE = 1.533 s (Table 5). No P- or S-phase arrivals were missed within the manually identified event windows. In the second event at station DKL, the model’s predicted S arrivals aligned more closely with the visually impulsive onsets than the manual reference pick. This suggests that the transformer-PE architecture captures coherent onset structures (Figure 10).

Station	Phase	Manual Time (UTC)	Error (s)	err (s)	Confidence
GADA	P ₁	2022-01-08 12:47:00.409	+2.629	2.629	0.968
GADA	S ₁	2022-01-08 12:47:17.229	+0.244	0.244	0.964
GADA	P ₂	2022-01-08 13:39:17.661	-1.720	1.720	0.961
GADA	S ₂	2022-01-08 13:39:21.975	-0.135	0.135	0.987
DKL	P ₁	2022-01-06 19:42:25.887	-0.025	0.025	0.983
DKL	S ₁	2022-01-06 19:42:31.126	-1.064	1.064	0.948
DKL	P ₂	2022-01-06 20:24:44.993	+0.928	0.928	0.987
DKL	S ₂	2022-01-06 20:24:52.648	-2.619	2.619	0.970

Table 5: Event-wise timing errors for all manually identified P and S arrivals in the 4-hour continuous test. P₁ and S₁ stand for P and S arrivals respectively for the first event in that station. For each arrival, all overlapping windows were evaluated and the smallest absolute error was selected (best-window error). Confidence refers to the softmax probability at the interpolated peak of the corresponding phase probability curve.

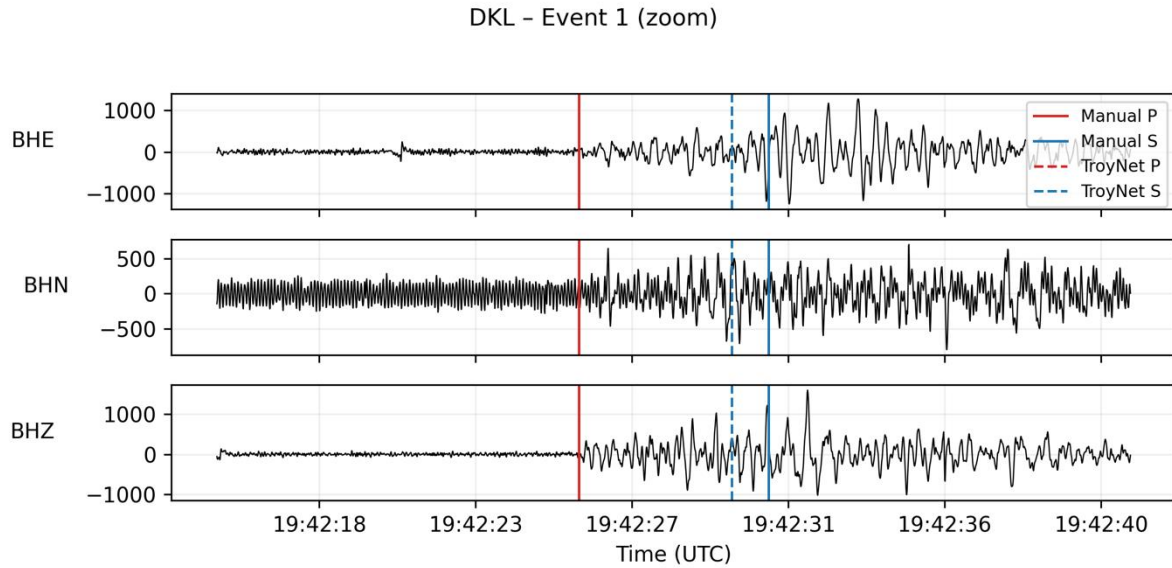


Figure 10: Zoomed view of Event 1 at station DKL. TroyNet-derived P and S picks closely match the manual analyst picks.

In addition to event-centered analysis, we also examined TroyNet's behaviour on noise-only windows. As expected for a pure phase picker trained exclusively on event windows, the model produced high softmax peaks in most noise segments, resulting in some false-positive picks (primarily for S phases) when applied without a preceding detection stage. True negatives indicate noise windows where no high-confidence phase pick is found. TroyNet is more permissive for S phases, reflecting their emergent nature. Phase-specific differences in noise behaviour are reflected in the false-positive statistics reported in Table 6. These results describe phase selectivity rather than event-detection capability, but also indicate that the P-phase false-positive and true-negative rates are favourable under noise-only conditions. Further study is required to analyse this behaviour during detection component development.

Metric	P-phase	S-phase
Noise windows	1420	1420
False positives	3	247
True negatives	1417	1173
False positive rate	0.002	0.174
False positives per hour	0.76	62.62

Table 6: Noise-window behavior on continuous data. False positives are counted on windows without manual arrivals and are reported both as proportions and as rates (FP/hour), computed using a 10 s sliding-window hop with each overlapping analysis window treated independently. Phase-specific thresholds are applied ($P \geq 0.97$, $S \geq 0.95$).

This behaviour does not reflect a deficiency of the picking model but rather the absence of a detection component. In operational settings, TroyNet should therefore be paired with an external trigger (e.g., STA/LTA, kurtosis, template detectors, or deep neural detectors) that first identifies candidate event intervals, after which TroyNet refines the P and S onsets. This aligns with standard practice in automatic processing pipelines. A detection component compatible with TroyNet is under investigation and will be addressed in the subsequent version.

3.4 Catalog Performance (Velocity Model & Epicenters)

As a final test, a randomly selected subset of 500 earthquakes from the test dataset was used to derive an optimal 1-D crustal velocity model. Each earthquake contained manually picked P- and S-phase arrivals recorded by at least four stations. Using this reference set, two additional earthquake subsets were prepared based on TroyNet- and PhaseNet-derived P- and S-phase picks. Again, the same requirement of recordings from a minimum of four stations was applied. As expected, the TroyNet-based dataset contained 395 locatable earthquakes, whereas the PhaseNet-based dataset contained 205 earthquakes, reflecting differences in S-phase detection performance. The initial velocity model was adopted with modifications from previous regional studies (Kalafat et al., 1987; Karabulut et al., 2003). Earthquake locations were computed using the HYP and VELEST subroutines of the SEISAN package (Havskov and Ottemöller, 1999). For VELEST, five iterations were performed for each pick set (manual, TroyNet, and PhaseNet) to ensure convergence. Figure 11 presents a comparison of RMS residuals obtained from epicentral solutions using HYP and VELEST for the three pick sources.

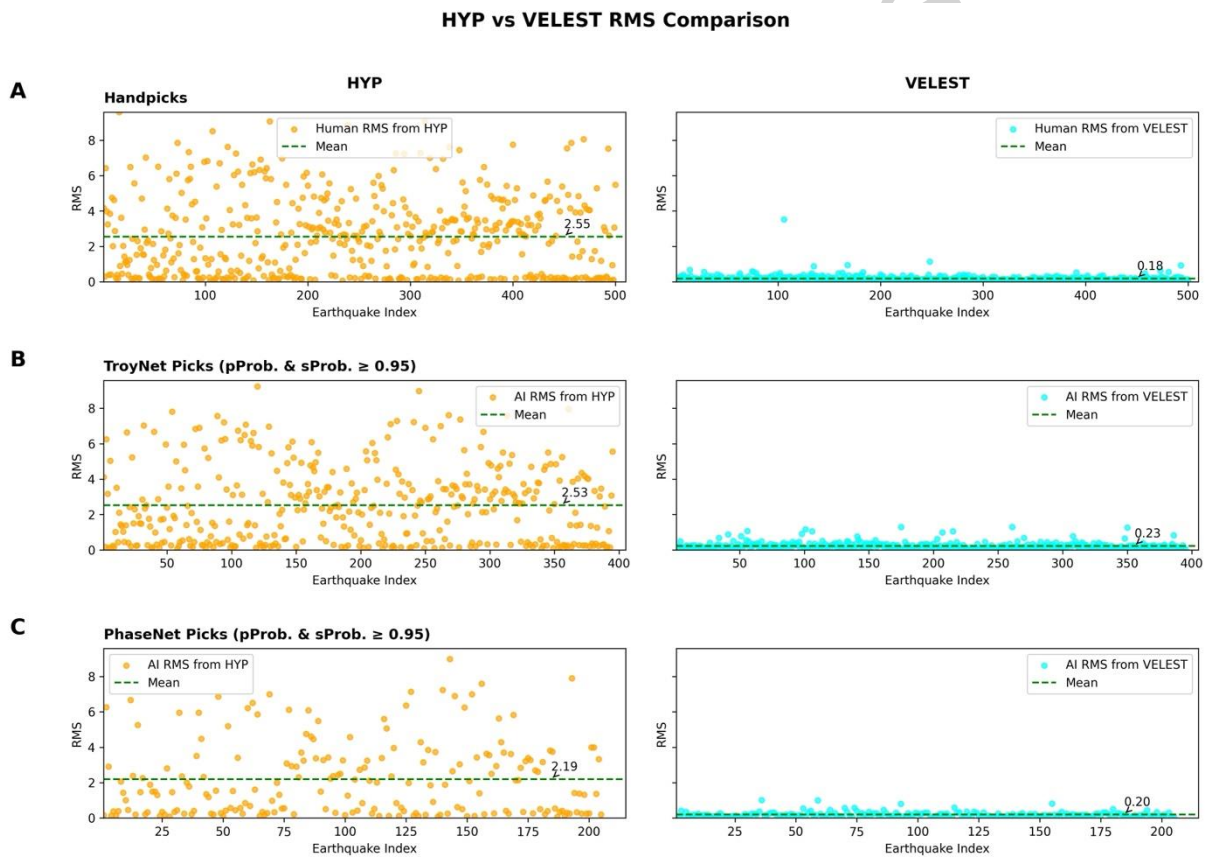


Figure 11: RMS residuals of epicentral solutions obtained using HYP and VELEST for manual picks (a), TroyNet picks (b), and PhaseNet picks (c). Only earthquakes with at least four stations were considered. VELEST solutions were computed using a common 1-D velocity model and five iterations. Dashed lines denote mean RMS values.

The results showed that VELEST consistently reduces RMS residuals relative to HYP for all pick sources. This confirmed the internal consistency of the derived 1-D velocity model and the stability of the inversion. As expected, locations based on manually picked arrivals yielded the lowest RMS values. RMS values obtained using TroyNet- and PhaseNet-derived

picks became comparable after inversion, hinting VELEST effectively mitigates moderate timing differences at the pick level. Both deep learning algorithms had similar post-inversion RMS performance. PhaseNet has slightly better RMS values but TroyNet enabled a larger number of earthquakes to be located compared to PhaseNet. This increase was primarily the result of TroyNet's higher P- and S-phase recall. It allowed more events to satisfy the minimum station and phase-pair requirements for reliable location. These results demonstrate that TroyNet improves catalog completeness without degrading location quality.

3.5 Sample Seismograms

Figure 12, 13, 14 and 15 present a representative subset of seismograms from the test set held-out during training, chosen to illustrate TroyNet's abilities over a range of signal-to-noise ratios and waveform complexities. Each panel shows one of the settings—from clean mid-SNR events to emergent or noisy arrivals—and compares TroyNet probabilistic picks with human catalog picks to illustrate both the model's strengths and its residual timing biases.

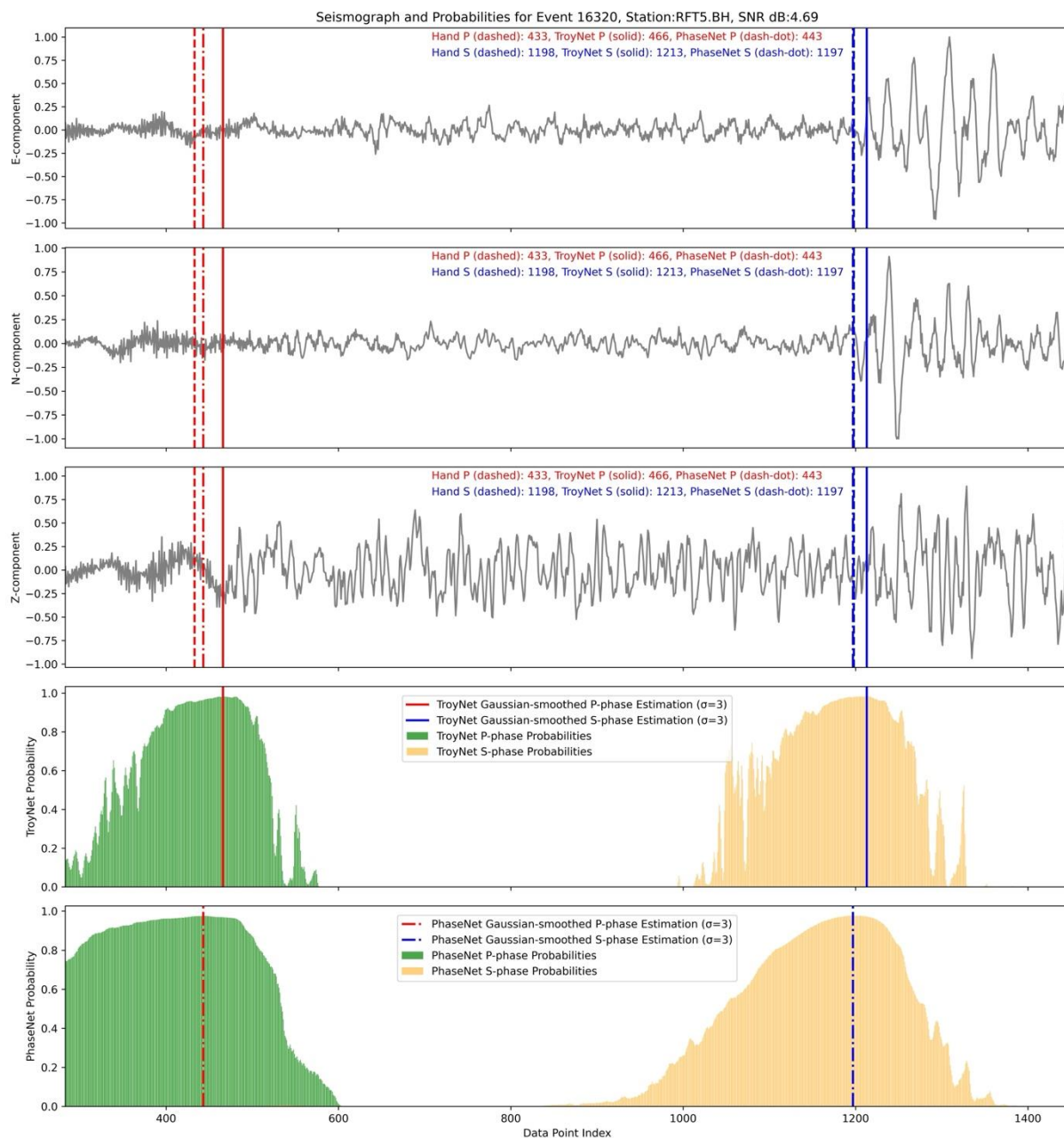


Figure 12 (Low SNR event): This figure shows a magnitude 2.2 earthquake recorded at station RFT5.BH on August 5, 2017. The top three panels display the E, N, and Z components of ground motion, and the bottom panel shows TroyNet's predicted P- and S-phase probabilities (green and blue, respectively). The manual picks are shown with dashed lines (red for P, orange for S), while TroyNet's predicted picks appear as solid lines. TroyNet makes confident picks with clearly separated P and S peaks, but both are slightly delayed compared with the manual picks, approximately 0.33 s for the P-phase and 0.15 s for the S-phase. This event is an example of how TroyNet can perform on low-SNR data with confident and well-placed picks.

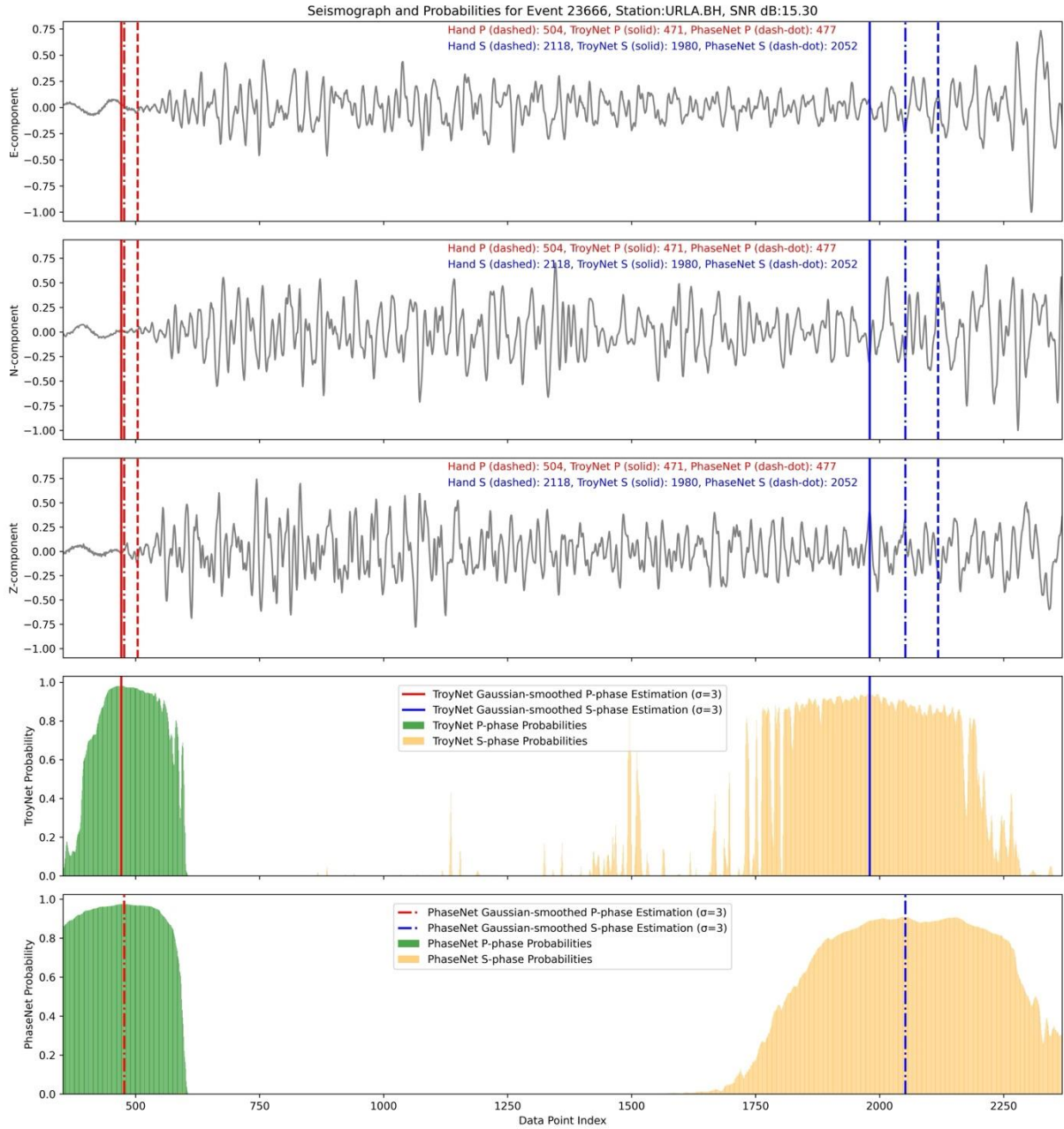


Figure 13 (Moderate SNR event): M 2.5 earthquake recorded at station URLA.BH on July 7, 2020. The top three panels display the E, N, and Z ground motion components, and the bottom panel shows TroyNet's P- and S-phase probability outputs. Manual picks are shown with dashed lines (red for P, orange for S), while TroyNet's picks are shown with solid lines. The P arrival (index 471) is detected by TroyNet with a residual of 0.33 s. Interestingly, the P pick made by TroyNet appears closer to the real onset of motion, especially for the Z component, which suggests that manual annotation might be somewhat late. The S arrival, clearest on the E component, emerges against a background of decaying P-wave energy. Since TroyNet is trained to detect the very first departure from background, tendency to pick the earlier onset is clear here. The analysts sometimes mark a slightly later point on emergent S arrivals to avoid the tail of the P coda.

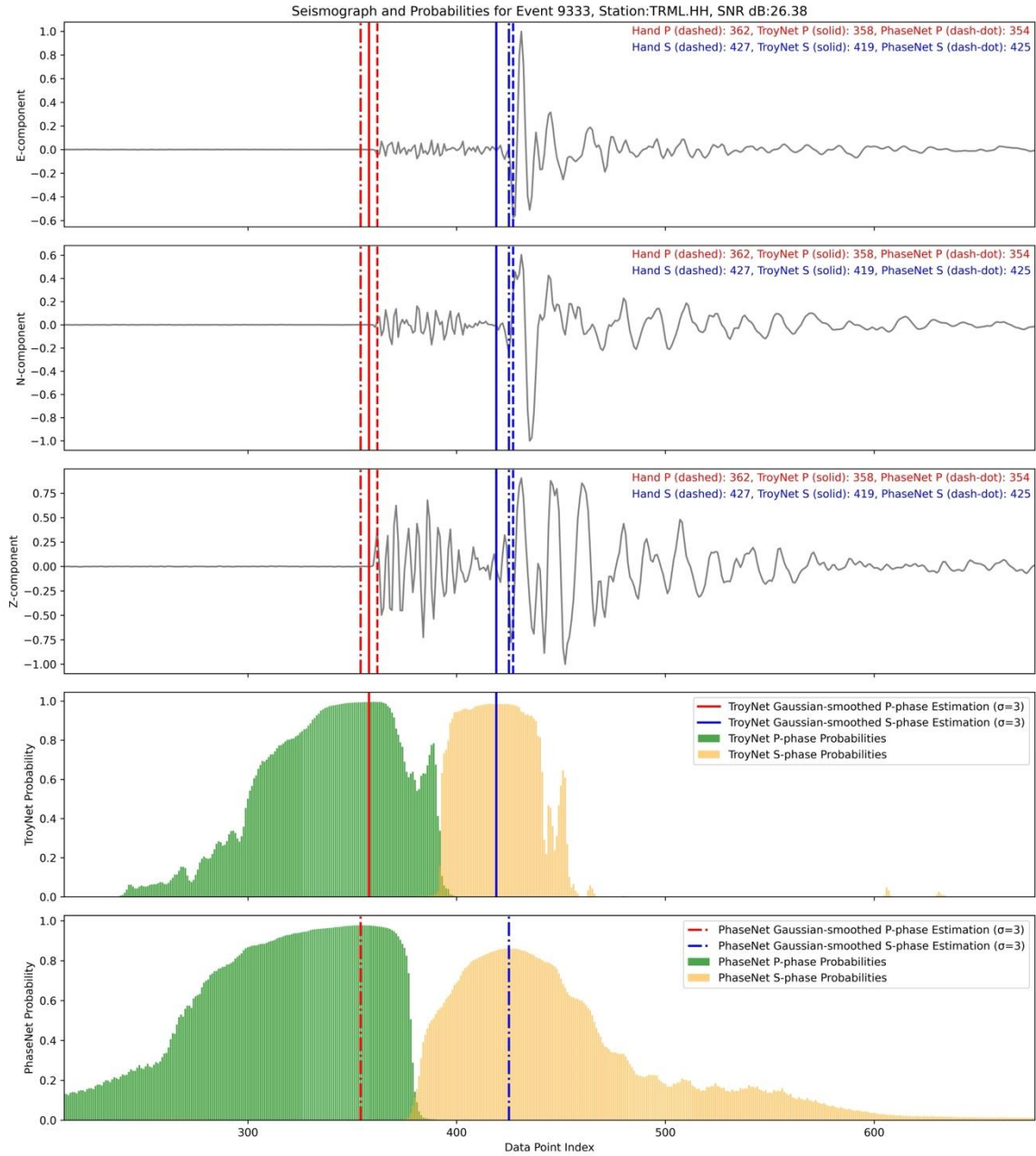


Figure 14 (High-SNR event): This plot shows a magnitude 1.2 earthquake recorded at the TRML.BH on March 18, 2015, with an RMSE-SNR (dB) of 26.38. The top three panels display the E, N, and Z ground motion components, and the bottom panel shows TroyNet's P- and S-phase probability outputs. Manual picks are shown with dashed lines (red for P, orange for S), while TroyNet's picks are shown with solid lines. TroyNet performs well in this high SNR case, producing both P and S picks with high confidence and very low residuals. The P-phase is picked at index 358, just 0.04 seconds before the manual label at 362. The S-phase is estimated at 419 versus a manual pick of 427 with 0.08 second difference. All three components show clear signal onset, and the probability curves for both phases are well defined and sharply peaked although the arrival times between phases are tight because of a small event.

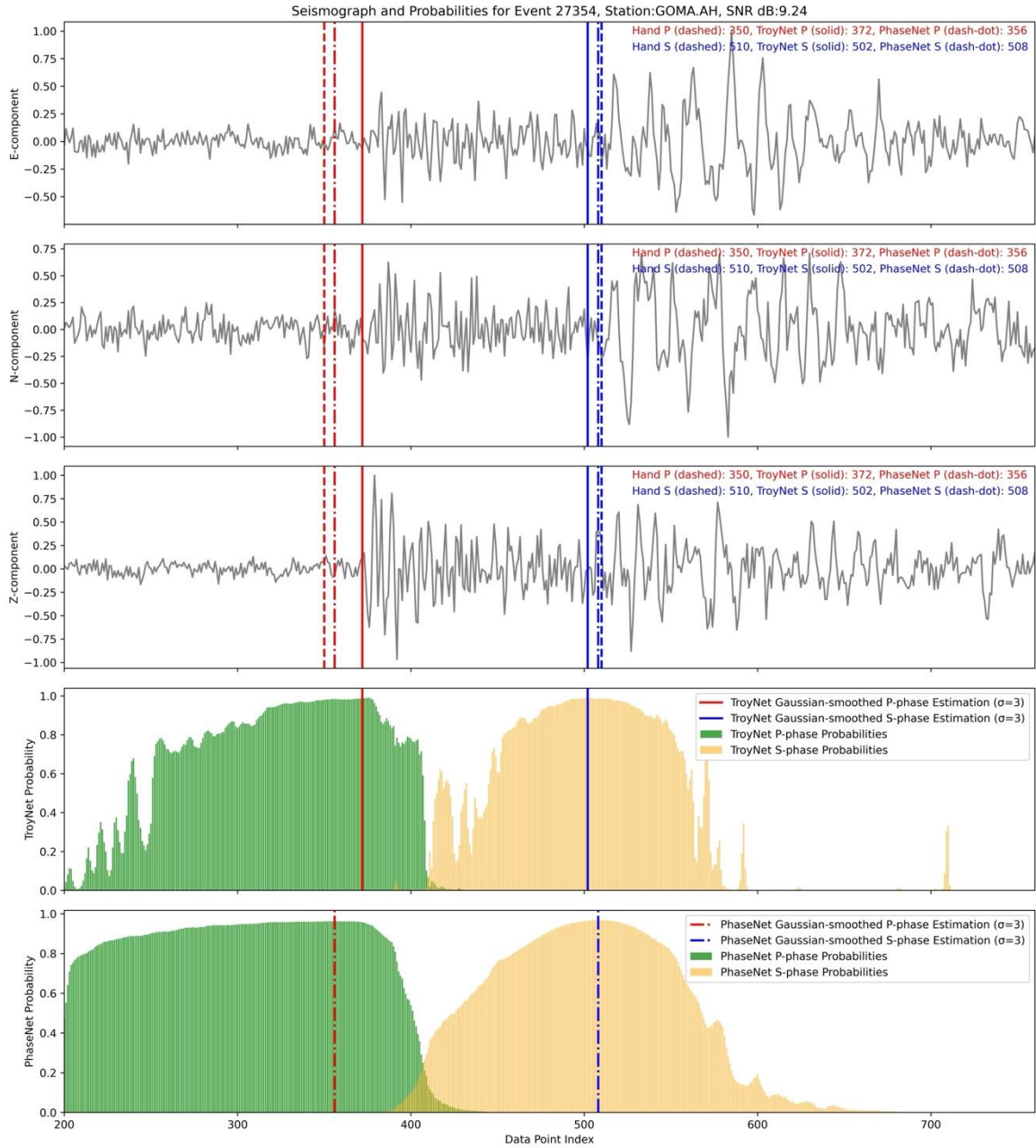


Figure 15 (Small magnitude event): This plot shows a magnitude 0.6 earthquake recorded at the GOMA.AH on December 7, 2020, with a low RMSE-SNR (dB) of 9.24. TroyNet identifies both the P and S arrivals with distinct probability peaks. The predicted P arrival is slightly delayed relative to the analyst pick (by ~ 0.22 s), whereas the S-phase estimate agrees closely with the manual reference (error ~ 0.08 s). The probability curves are broader and reflects the emergent character of the arrivals rather than a sharp impulsive onset.

4. Discussion & Conclusion

The deployment of TroyNet picker highlights several important points in practical seismic phase selection. The importance of regionally compatible training data is emphasized again. When PhaseNet was applied to northwestern Türkiye without local data training, it tended to make suboptimal choices because of variations in noise levels and possibly local wave propagation as shown by previous studies (Unal & Bekler, 2025). By being trained on local

waveforms, TroyNet learned these local characteristics, resulting in measurably improved performance for S-phase recall. This result is consistent with other research highlighting the importance of transfer learning or retraining for new regions.

When we put TroyNet head-to-head with retrained PhaseNet, TroyNet kept pace on P timing, lagged only slightly on S-arrival timing, yet pulled far more S-phase picks out of noisy records. Those gains are primarily associated with the transformer bottleneck, the U-Net-style skip links, and the positional encoding we introduced. In practice, that means a denser, more reliable earthquake catalogue for regional monitoring as results of continuous data test have indicated. If absolute precision with minimal false alarms is the only goal, for example in single-station tsunami warning, the tightened PhaseNet variant still has the edge. For day-to-day network operations, though, TroyNet strikes the better balance between accuracy, coverage and catalogue completeness. The continuous data test also pointed out that TroyNet is a good candidate as a picker, but it should be paired with a detection component that first identifies candidate event intervals, after which TroyNet refines the P and S onsets.

The use of an SGDR learning rate schedule was helpful for smoother convergence and reduced sensitivity to local minima. TroyNet's performance on noisy data has been shown to be effective, but other noise sources (such as cultural or seasonal noise) may require explicit handling. These factors were not explicitly modeled, yet the system may implicitly learn to distinguish them when trained on noise-only datasets.

TroyNet's early performance on event-based catalog data, which approaches that of human pickers, is an encouraging result for automated processing but it requires some key enhancements. Given TroyNet's competitive picking performance with additions of a positional-transformer encoder pair to acquire long-range temporal dependencies, dilated residual blocks for multiscale feature extraction, and a custom loss function to address extreme class imbalance, further architectural improvements will be investigated to reduce the residual timing error. One change is adding a regression head parallel to the classification output so that the model can estimate the timing explicitly as a continuum. Hybridization in this fashion may increase subsample accuracy and decrease bias, especially for emergent or low-amplitude arrivals. SNR-based training has been dropped from the workflow after the ablation study, but the results have shown that it is worth investigating further using alternative SNR-based learning strategies. Moreover, training and testing will be extended to other seismotectonic provinces in Türkiye, such as areas with deeper earthquakes, varying attenuation structures, or differing noise conditions. This will allow us to evaluate TroyNet applicability and direct the development of region-dependent modifications where needed. TroyNet processes each station's waveform independently. Future work will also include testing TroyNet with an earthquake detection mechanism on longer continuous data streams and evaluating catalog-level impacts through standard phase association workflows. Station-level and spatial performance variability, which reflects both picker behavior and site-specific effects, will be addressed in a dedicated follow-up study focused on operational deployment. These operational assessments are recognized as important for deployment but are beyond the scope of this methods-focused study, which centers on picker model development and event-based validation. With mentioned enhancements in mind, this study has shown that noisy seismic networks, such as those in Türkiye, can be processed in near-real-time with machine learning pickers, when combined with an external detection stage, to generate reliable phase arrivals for earthquake early warning or rapid response purposes.

References

1. Aguilar Suarez, A. L., & Beroza, G. C. (2025). Picking regional seismic phase arrival times with deep learning. *Seismica*, 4, e1431. <https://doi.org/10.26443/seismica.v4i1.1431>
2. Akol, B., & Bekler, T. (2013). Assessment of the statistical earthquake hazard parameters for NW Turkey. *Natural Hazards*, 68, 837–853. <https://doi.org/10.1007/s11069-013-0659-1>
3. Bekler, T., & Gürbüz, C. (2008). Insight into the crustal structure of the Eastern Marmara Region, NW Turkey. *Pure and Applied Geophysics*, 165, 295–309. <https://doi.org/10.1007/s00024-008-0302-3>
4. Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106, 249–259. <https://doi.org/10.1016/j.neunet.2018.07.011>
5. Caredda, E., Isken, M. P., Cesca, S., Errico, M., Zerbinato, G., & Morelli, A. (2025). Improving microearthquake detection in the Val d’Agri region (Southern Italy) with deep learning. *Seismica*, 4(2). Retrieved from <https://seismica.library.mcgill.ca/article/view/1690>
6. Büyükakpınar, P., Isken, M. P., Heimann, S., Dahm, T., Kühn, D., Starke, J., et al. (2025). Understanding the seismic signature of transtensional opening in the Reykjanes Peninsula Rift Zone, SW Iceland. *Journal of Geophysical Research: Solid Earth*, 130, e2024JB029566. <https://doi.org/10.1029/2024JB029566>
7. Havskov, J., and Ottemoller, L (1999). SeisAn Earthquake analysis software, *Seismological Research Letters*, 70, 532-534, doi:10.1785/GSSRL.70.5.532
8. He, K., Zhang, X., Ren, S., & Sun, J. (2019). Bag of Tricks for Image Classification with Convolutional Neural Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 558–567. <https://doi.org/10.1109/CVPRW.2019.00081>
9. Heuel, J., Maurer, V., Frietsch, M., & Rietbrock, A. (2025). Picking Induced Seismicity with Deep Learning (piSDL). *Seismica*, 4(2). <https://doi.org/10.26443/seismica.v4i2.1579>
10. Isken, M.P., Karstens, J., Nomikou, P. et al. Volcanic crisis reveals coupled magma system at Santorini and Kolumbo. *Nature* 645, 939–945 (2025). <https://doi.org/10.1038/s41586-025-09525-7>
11. Isken, M., Niemz, P., Münchmeyer, J., Büyükakpınar, P., Heimann, S., Cesca, S., Vasyura-Bathke, H., & Dahm, T. (2025). Qseek: A data-driven Framework for Automated Earthquake Detection, Localization and Characterization. *Seismica*, 4(1). <https://doi.org/10.26443/seismica.v4i1.1283>
12. Jamalreyhani, M., Wang, R., Schmandt, B., Felipe Peña Castro, A., & Glasgow, M. E. (2025). Evidence for fluid pressurization of fault zones and persistent sensitivity to injection rate beneath the Raton Basin. *Geophysical Research Letters*, 52, e2025GL114675. <https://doi.org/10.1029/2025GL114675>
13. Kalafat, D.; Gürbüz, C.; Üçer, S. B. (1987). Investigation of crust and upper mantle structure in western Turkey, *Bulletin of Earthquake Research Institute*, 59, 43-64
14. Karabulut, H.; Özalaybey, S.; Taymaz, T.; Aktar, M.; Selvi, O.; Kocaoğlu A. (2003). Tomographic image of the shallow crustal structure in the Eastern Marmara, *Geophysical Research Letters*, 30, 24, 2277, doi:10.1029/2003GL018074
15. Karabulut, H., Eken, T., Chen, Y., Taymaz, T., Houseman, G. A., & Saygin, E. (2024). The Marmara Sea basin as a regional depression constrained by seismological and geodetic observations. *Geophysical Journal International*, 237(2), 1208–1221. <https://doi.org/10.1093/gji/ggae109>

16. King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis*, 9, 137–163. <https://doi.org/10.1093/oxfordjournals.pan.a004868>
17. Loshchilov, I., & Hutter, F. (2017). SGDR: Stochastic gradient descent with warm restarts. *In International Conference on Learning Representations (ICLR)*. <https://arxiv.org/abs/1608.03983>
18. Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020). Earthquake transformer—An attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, 11, 3952. <https://doi.org/10.1038/s41467-020-17591-w>
19. Münchmeyer, J., Woollam, J., Rietbrock, A., Tilmann, F., Lange, D., et al. (2022). Which picker fits my data? A quantitative evaluation of deep learning-based seismic pickers. *Journal of Geophysical Research: Solid Earth*, 127, e2021JB023499. <https://doi.org/10.1029/2021JB023499>
20. Ross, Z. E., Meier, M.-A., Hauksson, E., & Heaton, T. H. (2018). Generalized seismic phase detection with deep learning. *Bulletin of the Seismological Society of America*, 108, 2894–2901. <https://doi.org/10.1785/0120180080>
21. Saad, O. M., Chen, Y., Siervo, D., Zhang, F., Savvaidis, A., et al. (2023). EQCCT: A production-ready earthquake detection and phase-picking method using the compact convolutional transformer. *IEEE Transactions on Geoscience and Remote Sensing*, 61, Article 4507015. <https://doi.org/10.1109/TGRS.2023.3319440>
22. Unal, U., & Bekler, T. (2025). Detection of P and S wave phases by machine learning using north-western Türkiye local seismic network data. *Pure and Applied Geophysics*, 182, 1381–1395. <https://doi.org/10.1007/s00024-024-03636-4>
23. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008. <https://arxiv.org/abs/1706.03762>
24. Xu, S., Wu, X., Li, Z., Wang, S., & Zhu, J. (2024). An all-in-one seismic phase picking, location, and association network for multi-station earthquake monitoring. *Communications Earth & Environment*, 5, 22. <https://doi.org/10.1038/s43247-023-01188-4>
25. Zhu, W., & Beroza, G. C. (2019). PhaseNet: A deep neural network-based seismic arrival-time picking method. *Geophysical Journal International*, 216, 261–273. <https://doi.org/10.1093/gji/ggy423>

Statement & Declarations

- **Funding:** The authors declare that no funding was received during the preparation of this manuscript.
- **Competing Interests:** The contact author has declared that none of the authors has any competing interests.
- **Author Contributions:** Conceptualization: U.U., T.B.; Methodology: U.U., T.B.; Software: U.U.; Validation: U.U.; Formal analysis: U.U.; Investigation: U.U.; Data curation: U.U.; Resources: U.U., T.B.; Writing – original draft: U.U.; Writing – review & editing: U.U., T.B.; Visualization: U.U., T.B.; Supervision: T.B.; Project administration: T.B.

Data Availability

A small sample of the actual test dataset and a test script are available in the Zenodo repository, <https://doi.org/10.5281/zenodo.15516676>

Accepted manuscript