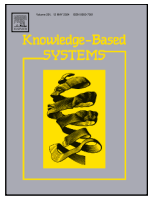
Knowledge-Based Systems 331 (2025) 114732

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys



Estimation of atmospheric visibility by deep learning model using multimodal dataset

Jitka Kopecká [a], Dušan Kopecký [a,b], Dominik Štursa [b], Zuzana Rácová [a], Tomáš Krejčí [b], Petr Doležel [b,*]

[a] Faculty of Chemical Engineering, University of Chemistry and Technology Prague, Technická 5, Prague 6, 16628, Czech Republic
[b] Faculty of Electrical Engineering and Informatics, University of Pardubice, Studentská 95, Pardubice, 53210, Czech Republic

ARTICLE INFO

Keywords:
Atmospheric visibility
Neural network
Multimodal dataset
Meteorological variables
Deep learning

ABSTRACT

Accurate estimation of atmospheric visibility is essential for numerous safety-critical applications, particularly in the field of transportation. In this study, a deep learning-based approach is investigated using a multimodal input representation that combines RGB images from a fixed-position surveillance camera with tabular meteorological variables collected from a nearby meteorological station. The meteorological input includes temperature, absolute pressure, relative humidity, dew point, wet bulb temperature, average and maximum wind speed, amount of precipitation, solar radiation, and ultraviolet index.

Six neural network models for visibility estimation were developed and compared: a multimodal model utilizing both image and tabular meteorological inputs; two ablation models that use only unimodal input (image or meteorological data); a regions-of-interest (ROIs) based model that extracts features from predefined image subregions; and two ablation models that use only a reduced number of meteorological data. The multimodal model uses EfficientNetV2M for feature extraction and a set of fully connected neural networks to integrate the two modalities. The ROIs-based model also uses EfficientNetV2M, but only on manually selected reference regions of the scene.

Evaluation was performed on a dataset of 1000 annotated images, with visibility manually determined based on reference points in the scene. The multimodal model achieved a mean squared error of 129,716 m^2, a mean absolute error of 165.4 m, and an R^2 score of 0.8861, with 84.46 % of predictions falling within a 10 % relative error margin. Although the ROIs-based model slightly outperformed the multimodal model in some regression metrics, its accuracy within tolerance thresholds was lower, and its reliance on manual scene annotation limits scalability. In contrast, the ablation models clearly demonstrated lower performance in almost all evaluated criteria.

The results display that the proposed multimodal input strategy provides a balanced and practical approach to automated visibility estimation. Compared to conventional unimodal input models, this architecture offers improved accuracy, stability, and generalisation ability, making it suitable for real-world applications where both visual and environmental data are available.

1. Introduction

Atmospheric visibility is a significant meteorological variable that is crucial to better and safer control of air, road, or sea transportation. Knowledge of the visibility distance, together with the possibility of predicting it in real time, especially during visibility degradation, allows one to increase operational efficiency, predict potential delays and avoid associated costs. Kim and Lee [1] Its knowledge is also essential for many other applications, for example the safe operation of autonomous vehicles [2] or planning the location of solar energy systems. Ekici and Teke [3] Visibility (by day) is the longest distance at which a black or dark object of appropriate dimension located on the earth's ground can be seen and recognized against the sky of the horizon during the day. World Meteorological Organization [4] The degradation of visibility can be related to both natural phenomena (such as rain, snow, fog, dust particles) and anthropogenic phenomena (such as solid or gaseous pollutants). The degradation of visibility is a frequent problem in urban regions of the developed world, where a high number of airports and a wide and dense network of roads are often found for which visibility knowledge and predictability are significant.

* Corresponding author.
E-mail address: petr.dolezel@upce.cz (P. Doležel).

https://doi.org/10.1016/j.knosys.2025.114732
Received 11 June 2025; Received in revised form 5 October 2025; Accepted 20 October 2025
Available online 26 October 2025

It is important to note that visibility degradation is strongly associated with increased concentrations of particulate matter (PM_{10}, $PM_{2.5}$) in the atmosphere, which also negatively affects human health. Specifically, high PM concentrations are a problem that mainly faces large Asian cities. Perhaps that is why extensive datasets monitoring environmental quality are being created there and subsequently analysed by a number of research groups. Chen et al. [5] If visibility prediction is the goal, it is therefore necessary to consider, in addition to meteorological conditions, also the sources that emit this type of pollution.

The simplest way to determine visibility is through observation and subsequent estimation by a trained human observer. The quality of such an estimate depends on the photometric character and dimension of the observed object, but also on subjective perception. World Meteorological Organization [4] Conventional measuring instruments are based on the interaction of electromagnetic radiation with particles present in the atmosphere. Typical instruments used at the airports are optical visiometres - transmissometres and forward scatter metres. However, these instruments suffer from a number of shortcomings, such as short measurement range, low accuracy for visibilities below 50 metres, or limited ability to provide real-time data. Chaabani et al. [6] Therefore, this raises the question of whether the visibility issue cannot be solved using computational methods without the presence of expensive single purpose measuring instruments.

The modern, but often experimental approaches are based on combination of cheap hardware with long-range visibility and/or real-time monitoring ability (e.g., sensor modules, RGB cameras) and sophisticated computational algorithm primarily utilizing unimodal input, i.e., tabular or image datasets. The tabular datasets include information on meteorological conditions (such as temperature, surface wind speed, precipitation, or particulate matter concentration) [7,8] often obtained by professional synoptic, climatological, or airport meteorological stations. However, image datasets gained from camera-based methods, random city cameras [9] or even remote sensing satellites [10,11] have recently become the subject of interest. Liu et al. [12], Gáborčíková et al. [13] An indisputable advantage of cameras is the significantly lower costs of purchasing and operating compared to expensive specialized visiometres. A major disadvantage of image datasets can be seen in insufficiently large, inconsistent, and sometimes even synthetic-created images. There are also computational methods available that use multimodal input datasets. However, this often involves double use of the same image, the original and its pre-processed version, e.g., pseudo-colour image [12] or infrared image Wang et al. [14].

Different computational approaches are used to determine visibility. The first studies derived visibility from the calculation of extinction coefficient using meteorological laws, i.e., the Koschmieder equation and the Lambert-Beer law. Babari et al. [15], Carretas et al. [16] Later, research focused on the use of statistical methods, such as integrated autoregressive moving average [7,8] or linear and multiple regression models Sabetghadam and Ahmadi-Givi [17].

With the advancement of machine learning and neural networks (NN), these tools are also being tested on meteorological issues. Liu et al. [18] Handcrafted or learnt feature extraction is a common procedure that is closely associated with visibility estimation. The review article by Ouadil et al. [19] summarizes in detail the current state of knowledge on visibility estimation using NN. Among the most frequently tested methods of machine learning are support vector machines [7,20], random forest [7], K-means Majewski et al. [21], Sládek and Donateo [22].

It is valuable to conduct a detailed analysis of the use of artificial NN for visibility estimation. An effective tool for the input image dataset is the convolutional neural network (CNN), which is a feed-forward NN that uses convolution to automatically extract features from images. Li et al. [9], Gáborčíková et al. [13], Chaabani et al. [23], Rejfek et al. [24] Recurrent neural network (RNN) offers a different approach, processing sequential data by maintaining an internal memory, allowing it to capture temporal dependencies and patterns in tasks such as language modelling and signal processing. Wen et al. [7], Chen et al. [25], Chu et al. [26] Moreover, the transformer-based approach employs a sequence-to-sequence architecture with an encoder-decoder framework, where the encoder extracts features from the input sentence, which are subsequently utilised by the decoder to generate the output sentence. Its modification Visio Transformer is used for image processing and subsequent visibility estimation. Liu et al. [27,28] Some research integrates multiple methods simultaneously to enhance accuracy, such as the CNN-RNN combination. Song et al. [29] The article [20] uses the CNN-RNN model for visibility estimation, but the input image dataset is based on simply comparing the visibility level of a pair of images.

There are several challenges directly connected with area of visibility estimation using deep learning methods. The first challenge is related to the limitations that arise from the modality of the data. Unimodal models struggle under low contrast, backlight/direct sun light, snow/rain or sensor artefacts (image models), or they cannot handle scene specific issues, e.g., haze, skyline contrast (tabular meteorological data models). The second challenge is the dependence on hand-picked areas (ROIs) or scene calibration, which results in poor transfer possibilities. There are also limited possibilities to obtain public, real-word multimodal datasets, especially from a region of Europe. The reported metrics (MSE / MAE) are also rarely aligned with operations as with the classical metrics $\pm 10\,\%/\pm 20\,\%$ within-tolerance rates, directly reflecting decision thresholds.

In this study, the multimodal input dataset is used for visibility estimation, that is, fusion of images and tabular meteorological variables. Multimodal dataset for visibility estimation was recently used by Chen et al. [25]; multimodal fusion was performed using meteorological data and manual extraction of image features for RNN. Moreover, recent advances in multimodal deep learning cover CNN-based mid-level fusion for multimodal remote sensing classification, demonstrating the benefit of learned cross-modal interactions [30], parameter-efficient adapters on top of foundation models such as SAM to encode invariances and improve transfer in complex urban scenes [31], and fusion-aware transformers that exploit spectral-aware self-attention and cross-attention for computational imaging. [32] Herein presented study differs in scope (scalar visibility regression from low-cost camera and meteorological data), but leverages the same principle of shared latent space fusion while prioritizing deployability on modest datasets. Therefore, this work adopts a compact CNN backbone with a fully connected NN meteorology data branch and discusses adapter-style conditioning and cross-attention between image and meteorological data as future extensions.

Data for both modalities are collected from the university meteorological station for a period of one year, which allowed us to encounter all the common meteorological phenomena occurring during the whole year (haze, rain, snow, different intensity of solar radiation, etc.). This extended collection period ensures that the dataset captures realistic seasonal variability and represents a robust basis for training and evaluation. Subsequently, 1000 images of the longest possible visibility and reduced visibility and the respective meteorological variables were selected to use as a data source for machine learning purposes.

The main contribution of this study lies in the design and evaluation of a deep neural model that effectively integrates image and tabular meteorological data in shared latent space for improved visibility estimation. A convolutional backbone (EfficientNetV2M [33]) is used for image processing, while meteorological inputs are handled via a dedicated fully connected network. The quality of the resulting predictions is evaluated using standard regression metrics, including Mean Squared Error, Mean Absolute Error, Coefficient of Determination, Linear Regression Fit, and Percentage Error within predefined tolerance thresholds. A direct comparison of metrics on an ablation study (images only, tabular data only, and fusion of images and meteorological data) is presented, and a detailed cross-validation study allows us to assess the overall contribution. Finally, from a practical point of view, a real multimodal dataset (visual records paired with co-located meteorological

Fig. 1. The typical high (top) and low (bottom) visibility observed by camera (8 am, maximal distance 4 km).

measurements) uploaded to Zenodo is released with a long range of visibility in the geographical area of Europe.

2. Data collection and annotation

2.1. Dataset of meteorological variables

Dataset of meteorological variables was obtained using a Vantage Pro2 Plus meteorological station (Davis Instruments, USA) located on the roof of Building B of the University of Chemistry and Technology, Prague, Czech Republic (50°06'07.5"N, 14°23'22.4"E; altitude 239 m above sea level). In total 41 measured meteorological variables and calculated indexes were stored every 15 m for the period of 1 year (04/2023 - 04/2024) in the Weatherlink.com cloud and later exported as tabular data into a csv file. More than 35 000 records consist of detailed information about local meteorological situation and cover the studied visibility experiment without a time interruption. Later, the meteorological variables were narrowed from 41 to 10 most important; see Section 2.3.

2.2. Dataset of visibility images

Images of atmospheric visibility were captured using a Foscam NVR kit of four cameras (Foscam, China) attached to the stable base of the meteorological station. Each camera was directed to a different cardinal direction, and, due to the high density of development and slightly rugged georelief, they observed different longest distances of the scene (from hundreds to several thousands of metres). Each camera recorded a one hour long video in four hour long intervals (4 am, 8 am, 12 pm, 4 pm, 8 pm). Data were transmitted over Wi-Fi to the main camera unit. Each video was stored on the hard drive of the main camera unit in 1080p HD format (H.265 encoding).

All collected videos were manually presorted into the group with the longest possible visibility and the group containing the videos with reduced visibility. All videos were subsequently cut into JPG images with a resolution of 1920 × 1080 at 5-m intervals. Finally, since the longest possible visibility observed by the camera (that is, a cloudless or sunny; visibility approximately 4 km) is a more common phenomenon than reduced visibility at this particular latitude, 30 % images with the longest visibility and 70 % images with reduced visibility were randomly selected from each group and a dataset containing 1000 images was created. For a typical example of visibility images, see Fig. 1.

The vertical relief of the terrain in the observed scene is shown in Fig. 2. The terrain is characterised by a ~1 km wide zone (the valley of the Vltava River) below the horizon of the camera, which lacks any suitable reference points. The distance in this dead zone cannot be determined reliably and therefore cannot be explicitly used for image annotation. This gap introduces an inherent limitation into the ground truth data, as it prevents differentiation of visibility values within this

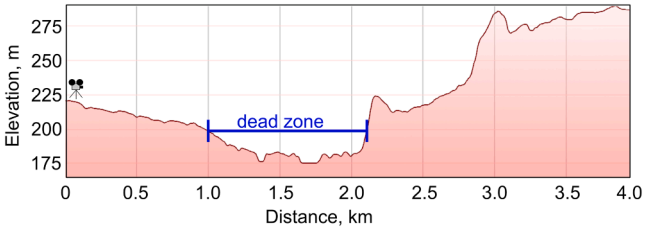


Fig. 2. The vertical relief of the terrain with a marked camera dead zone (adapted according to Google Earth).

Fig. 3. Reference points used for annotation of visibility images.

distance range. Similar discontinuities in the availability of reference objects have been reported in other visibility monitoring studies in complex urban or mountainous environments (e.g., [9,34]), which confirms that this is not an isolated issue of presented dataset but rather a general challenge for image-based visibility estimation. From a practical point of view, relatively short scenes or visually interrupted sight lines are common in densely populated urban areas, and any urban meteorological station recording image data must deal with such situations. In this sense, the presented dataset of visibility images still fits well into the expected real-world conditions of visibility measurement while also providing a clear direction for methodological improvements.

The experiment (training and initial testing) was also limited to a specific recording time of 8 am to reduce the impact of direct solar radiation on the cameras.

The distance of all reference points in the camera field of view, suitable for visibility evaluation, has been measured using Google Earth (Alphabet, USA) measuring tools and used for image annotation. The reference points used to determine visibility are shown in Fig. 3. The ground truth visibility was estimated by locating the furthest detectable reference point in the image. All images were annotated using this method by two independent observers using in total 53 reference points in the scene. The procedure corresponds to the real-world approach of meteorologists, who in practice estimate visibility by searching for known reference points in the landscape. Two histograms of the distances of the reference points in the scene used for image annotation and visibility in a dataset annotated by observers are shown in Fig. 4. Unlike the assessment of visibility in the open landscape of remote meteorological stations, where the distances between reference points are low, the assessment of visibility in an urban environment from a single observation point is usually associated with unevenly distributed reference points. Herein, a gap is created by the Vltava River valley, as observable in Fig. 4.

The dataset of visibility images thus provides a diverse collection of scenes under various atmospheric conditions. Each image is timestamped, which allows temporal alignment with the corresponding meteorological records.

2.3. Association of images with meteorological variables

For the purpose of multimodal learning, each visibility image was paired with the corresponding set of meteorological variables. The association was carried out on the basis of the closest matching timestamps between the camera acquisition and the meteorological station records. In cases where minor discrepancies occurred, linear

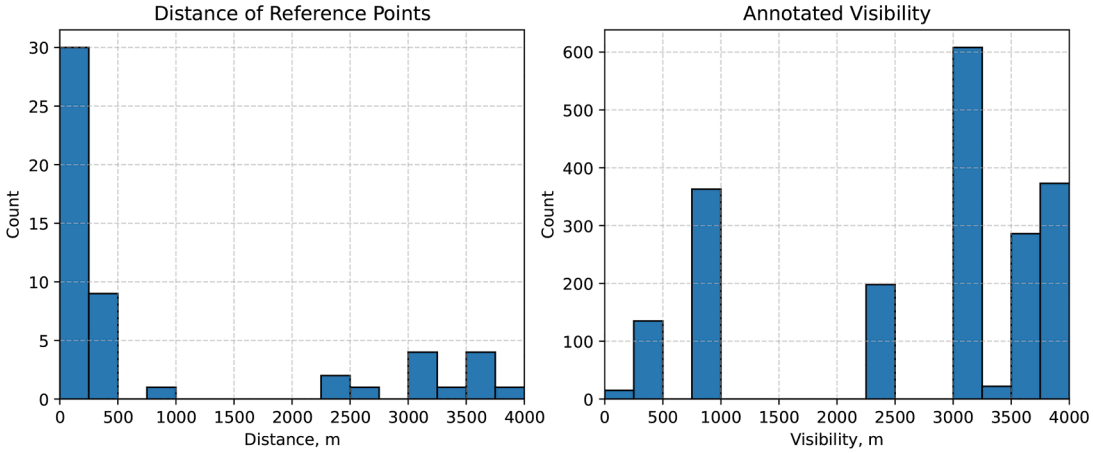


Fig. 4. Histogram of distances of reference points used for annotation in the scene and histogram of distances of visibility annotated by observers.

interpolation was applied to ensure temporal consistency. This procedure resulted in a synchronized multimodal dataset, where every image sample is complemented by a tabular vector of meteorological measurements.

From all the variables provided by the meteorological station (see Section 2.1), 10 relevant meteorological variables and calculated indexes were selected for NN training purposes. Specifically, temperature (unit degree Celsius) expressing the thermal state of the meteorological station's surroundings; absolute pressure (unit hectopascal) is measured relative to a perfect vacuum. Furthermore, variables quantifying atmospheric humidity as relative humidity (unit percentage), dew point (unit degree Celsius), and wet bulb temperature (unit degree Celsius). From this list, relative humidity is a ratio of the absolute humidity of the air to the absolute humidity of the air saturated by water vapour under given conditions (temperature, pressure). A dew point is the temperature to which humid air must be isobarically cooled to achieve its full saturation. Wet bulb temperature is the lowest temperature caused by the evaporation of water under given conditions (temperature, pressure).

Other variables include: the average wind speed (unit kilometres per hour) expresses the wind velocity as averaged over a period of 10 minutes. The maximum wind speed (unit kilometres per hour) is the maximum value of the wind speed in the last 10 minutes. The amount of precipitation (unit millimetre) represents the amount of water in a liquid or solid state that falls on a horizontal surface of the earth at a given location for a certain period of time. Solar radiation (unit watt per square metre) expresses global solar radiation (including direct and diffuse solar radiation). The ultraviolet (UV) index (dimensionless) expresses the intensity of UV radiation. Naturally, the influence of individual meteorological variables on the accuracy of visibility estimation may vary. Some features may provide highly complementary information, while others could be redundant or exhibit a weak correlation with actual visibility conditions. For instance, Liu et al. [12] identified temperature, absolute pressure, relative humidity, and average wind speed as the most influential meteorological predictors for visibility estimation tasks. This subset of tabular meteorological variables is therefore selected and evaluated in the ablation study.

2.4. Data preparation for machine learning

The annotated dataset comprising 1000 paired samples - each consisting of a visibility image, the corresponding vector of meteorological variables, and the annotated visibility - was randomly divided into three subsets for machine learning purposes. Specifically, 70 % of the data (700 samples) were used for training, 15 % was reserved for validation during model optimization, and the remaining 15 % formed the test set for the final evaluation.

This stratified division was chosen to ensure that each subset reflects the distribution of visibility levels and meteorological conditions present in the complete dataset. The distribution of visibility levels for each set is shown in Fig. 5.

The relatively large training portion ensures that the model has sufficient examples to learn complex multimodal relationships, while the separate validation and test sets enable robust performance tracking and protect against overfitting or overestimation of predictive accuracy.

To explicitly assess the influence of the dead zone (see Section 2.2) and the associated uncertainty in the ground truth, an additional dead-zone-filtered test set was prepared from the original test partition by removing all samples with annotated visibility in the range 750–2500 m. Samples within this interval coincide with the region where the absence of reference points prevents reliable discrimination of true visibility and, therefore, present a higher risk of annotation bias. The dead-zone-filtered test set represents approximately 10 % of the entire dataset, which still provides a statistically significant number of samples for evaluation.

The proposed models are evaluated on both the original (unfiltered) test set and the dead-zone-filtered test set. Performance in the latter is expected to provide a clearer estimate of the ability to predict visibility where the ground truth is reliable. Furthermore, a comparison between the two evaluations quantifies the sensitivity of the models to the potential bias introduced by ambiguous annotations in the 750–2500 m range. This test partition thus represents an explicit step towards accounting for uncertainties in the ground truth and assessing model robustness under realistic conditions in which visually interrupted sight lines commonly occur.

The complete dataset, structured and annotated as outlined in this section, is publicly available via the EU Open Research Repository ZENODO. It can be accessed at: doi:10.5281/zenodo.15494899.

3. Methodology

The goal of this study is to develop a robust deep learning model to estimate atmospheric visibility using a multimodal dataset. Unlike conventional methods that rely solely on image-based visibility estimation [9] or meteorological variables [8], the presented approach integrates both visual information from RGB images and meteorological variables from meteorological station. This multimodal fusion aims to improve predictive accuracy by incorporating atmospheric conditions that directly influence visibility.

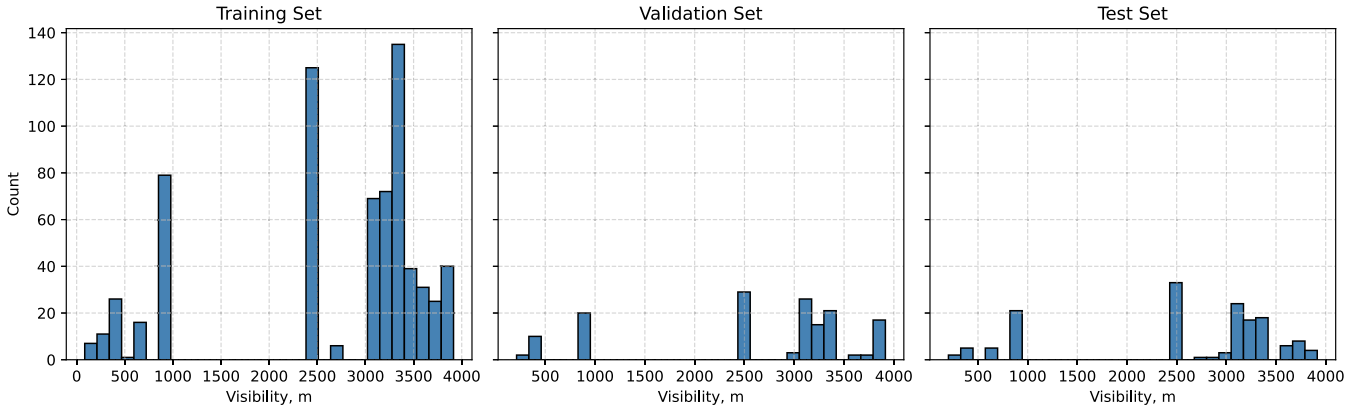


Fig. 5. Distribution of the visibility range in the training, validation, and test sets.

3.1. Multimodal data representation

To effectively model atmospheric visibility, the presented method processes two distinct data sources:

- RGB images captured from surveillance cameras provide spatial and contrast-based cues essential for estimating visibility. These images are processed using a CNN to extract high-level feature representations.
- A set of numerical meteorological variables collected from the meteorological station, including temperature, absolute pressure, relative humidity, dew point, wet bulb temperature, average wind speed, high wind speed, amount of precipitation, solar radiation, and UV index. These variables contribute crucial information on atmospheric conditions that affect visibility.

The challenge in combining these two modalities lies in their differing dimensionalities: image-derived feature vectors are high-dimensional, while tabular meteorological data consist of only a few numerical values. To address this, a neural architecture that balances both inputs in a unified latent space is designed.

3.2. Deep learning model architecture

The proposed NN consists of two processing branches that independently transform the two data modalities before merging them for the final prediction:

- Feature Extraction from Images: An EfficientNetV2M model [33], pre-trained on ImageNet, with its top layers removed. This CNN processes input images of size 224×224 and outputs a feature vector of dimension 1280.
- Transformation of Tabular Meteorological Data: A fully connected network maps the 10-dimensional tabular input into a higher-dimensional latent representation.

The network architecture consists of the following layers:

1. The 1280-dimensional feature vector extracted from EfficientNetV2M is passed through three dense layers (512, 256, and 128 neurons) with ReLU activation functions, followed by a dropout layer (rate = 0.1) to prevent overfitting.
2. The tabular meteorological data input (10-dimensional) is fed into a dense layer with 128 neurons and ReLU activation functions, followed by a dropout layer (rate = 0.1).
3. The outputs of both branches are concatenated and passed through additional dense layers (128 and 64 neurons) with ReLU activation functions, followed by a final dense layer with a single neuron and a linear activation function to produce the visibility estimate.

The overall architecture is depicted in Fig. 6 and is referred to in the text below as Multimodal (Image and Tabular) Model.

3.3. Training and optimization strategy

In order to effectively train the neural model proposed for atmospheric visibility estimation, a strategy was adopted that ensures robust generalisation while maintaining computational efficiency. Given the multimodal nature of the input data, which includes high-dimensional deep features extracted from images and low-dimensional tabular meteorological data, it is crucial to balance their influence during training. To achieve this, both data streams are pre-processed independently before merging them at later stages of the model. Furthermore, due to the inherent variability in atmospheric conditions and the potential for noise in both image and meteorological data, real-time data augmentation and normalization techniques were employed. These steps mitigate overfitting and enhance the ability of the model to generalise in various meteorological conditions. Additionally, since the dataset consists of sequentially acquired meteorological observations, a data shuffle is implemented at each epoch to prevent the model from learning temporal dependencies that may bias the predictions.

To train the model efficiently, a custom data generator is utilized. This generator:

- reads image filenames and corresponding tabular meteorological data from CSV files;
- applies real-time data augmentation (horizontal flips, small rotations, and shifts) to improve generalisation;
- normalizes meteorological input data and visibility target values by scaling them to the range $[0, 1]$ to stabilize training;
- extracts EfficientNetV2M features in batches, ensuring efficient GPU utilization.

A diagram of the custom data generator is shown in Fig. 7.

Training is carried out using the Adam optimizer with a learning rate of 0.001, optimizing the model with a loss function of mean squared error. The model is trained for up to 1500 epochs to ensure stability, with early stopping mechanisms governed by validation loss monitoring.

Separate data generators are instantiated for both training and validation, with shuffling enabled only during training to prevent overfitting. Randomly sampled 15 % of the dataset is used for validation. The model that performs best with the validation data is stored for further evaluation.

This structured approach ensures that both image and tabular meteorological data contribute meaningfully to visibility estimation, using deep learning to model complex relationships between meteorological conditions and observed visibility levels.

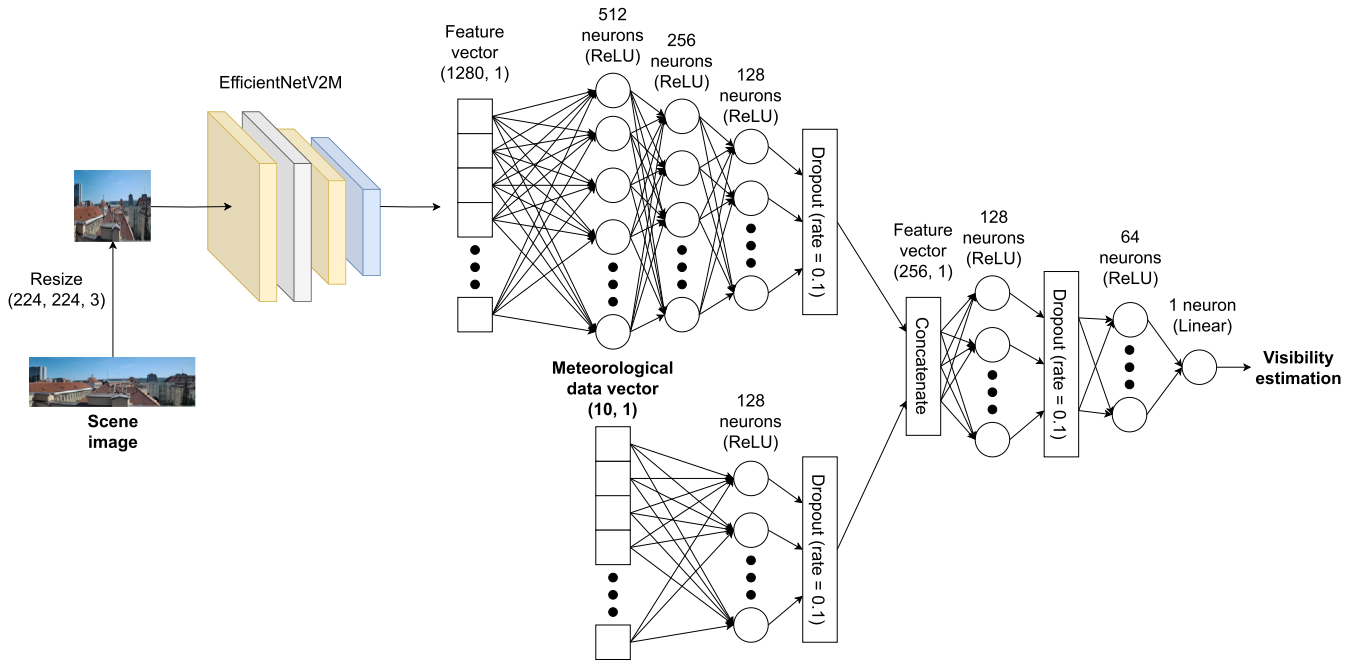


Fig. 6. Deep learning model architecture for visibility estimation from scene image and tabular meteorological data - Multimodal (Image and Tabular) Model.

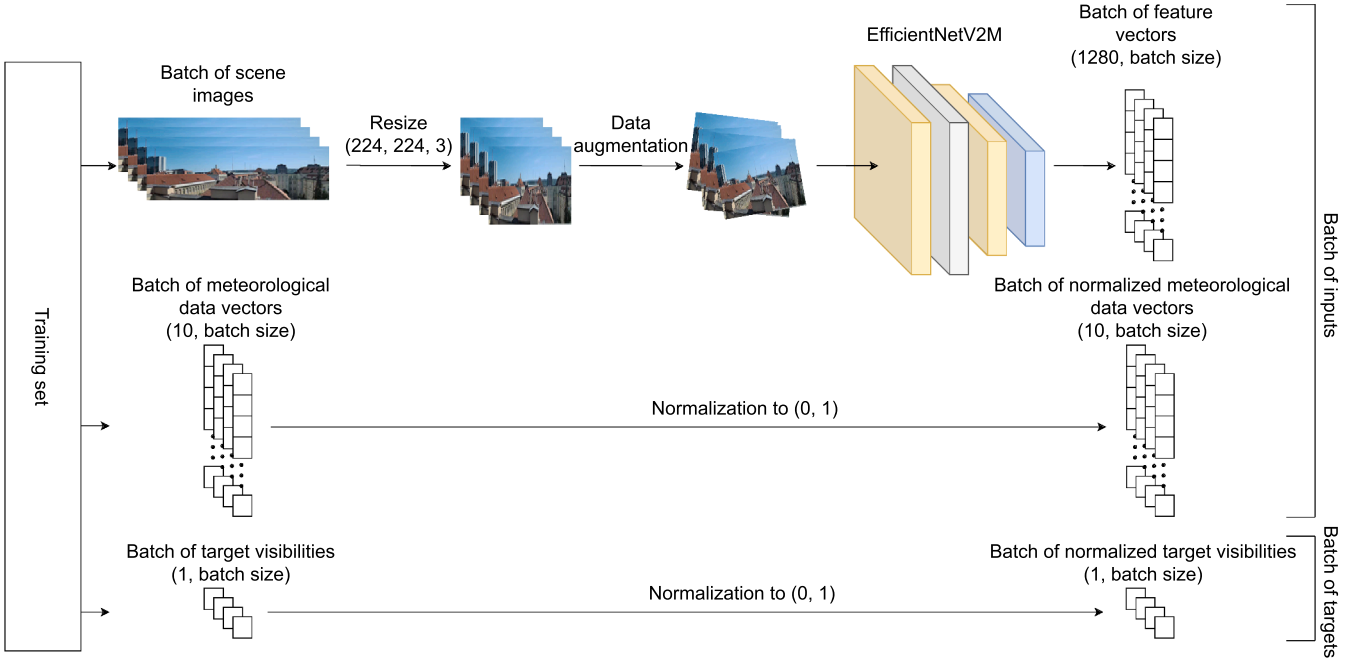


Fig. 7. Custom data generator to produce a batch of training inputs and training targets.

3.4. Metrics and results evaluation

To evaluate the performance of the proposed NN model in the test dataset, multiple regression-based metrics are employed to assess both absolute and relative errors, as well as the linearity of predictions. These metrics provide a comprehensive understanding of model accuracy, stability, and generalisation capability.

3.4.1. Mean squared error (MSE)

The primary evaluation metric, consistent with the loss function used during training, is the mean squared error (MSE), which is defined as:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2, \tag{1}$$

where y_i represents the ground truth visibility, $\hat{y}_i$ denotes the predicted visibility, and N is the number of test samples. Since large deviations are penalized more heavily than small ones, this metric is particularly sensitive to significant prediction errors.

3.4.2. Mean absolute error (MAE)

To complement MSE, the mean absolute error (MAE) is also computed:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|. \tag{2}$$

Unlike MSE, which squares errors, MAE treats all deviations linearly, providing a more interpretable measure of the average prediction error in physical units of visibility.

3.4.3. Coefficient of determination (R^2 score)

The coefficient of determination, also called R^2 score, is used to quantify how well the model explains the variance in the data:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}, \tag{3}$$

where $\bar{y}$ represents the mean of the observed values. An R^2 value close to 1 indicates that the predictions align well with the ground truth, while a value close to 0 suggests weak predictive performance.

3.4.4. Linear regression fit

To analyse the correlation between predictions and ground truth, a linear regression model is of the form:

$$\hat{y} = ay + b. \tag{4}$$

The regression model is fitted using least squares estimation, where a and b are the regression coefficients. Ideally, a slope a close to 1 and an intercept b near 0 indicate minimal systematic bias in the predictions.

3.4.5. Percentage error within a tolerance threshold

For practical applications, the proportion of predictions falling within a given percentage error margin is computed:

$$\mathrm{PE}_{\delta} = \frac{1}{N}\sum_{i=1}^{N}\mathbb{1}\left(\left|\frac{y_i - \hat{y}_i}{y_i}\right| < \delta\right), \tag{5}$$

where $\mathbb{1}(\cdot)$ represents the indicator function, and δ denotes the tolerance threshold (e.g., 10 % or 20 %). This metric reflects the reliability of the model in real-world scenarios where small errors are considered acceptable.

3.5. Competitive approaches

To comprehensively assess the effectiveness of the proposed multimodal approach, comparative analyses with alternative methodologies are performed. These include an ablation study to isolate the contributions of individual data modalities and an approach that uses regions-of-interest (ROIs) selection for feature extraction.

3.5.1. Ablation study

To assess the impact of combining image and tabular meteorological data, four ablation experiments are performed to isolate the contributions of individual modalities and variable subsets.

- Unimodal (Image) Model: In this variant, only the extracted deep features from the EfficientNetV2M model are used as input. The tabular meteorological data input branch is removed and the visibility estimation relies solely on spatial and contrast-based cues present in the image. This setup allows for the evaluation of how much visual information alone contributes to the final estimation. The visualisation of this competitive approach is shown in Fig. 8.
- Unimodal (Tabular) Model: In this configuration, only tabular meteorological variables are used for the estimation, while CNN-based image feature extraction is omitted. This approach assesses the extent to which numerical meteorological attributes can estimate visibility without the aid of visual cues. The visualisation of this competitive approach is shown in Fig. 9.
- Unimodal (Tabular - reduced) Model: This model is structurally identical to the Unimodal (Tabular) Model but uses only a selected subset of four meteorological variables: temperature, absolute pressure, atmospheric humidity, and average wind speed. This subset was chosen based on [12] as the most influential variables on visibility conditions. By comparing this model to the full Unimodal (Tabular) variant, the effect of reducing input dimensionality to only the most informative features can be evaluated.
- Multimodal (Image and Tabular - reduced) Model: This variant extends the concept of input reduction to the multimodal setting. It combines the full RGB scene image with only the aforementioned four key meteorological variables. The tabular branch is structurally unchanged but accepts a 4-dimensional input vector instead of the full 10-dimensional one. This model allows assessing whether the multimodal fusion benefits are preserved even when the meteorological component is simplified.

By comparing the performance of these ablated models to the complete Multimodal (Image and Tabular) Model, the contribution of each data source and the robustness of the proposed multimodal architecture with respect to input complexity are systematically evaluated. The combined model is expected to outperform both individual approaches by capturing a more comprehensive representation of atmospheric conditions. In addition, reduced models can provide insight into whether a complete set of meteorological variables is necessary for reliable visibility estimation.

3.5.2. Regions-of-interest-based feature extraction

Another competitive approach involves refining the image-based feature extraction by utilizing only specific ROIs rather than the entire scene. These ROIs correspond to manually identified reference points in the image that were used to tag visibility distances (see Fig. 3). The selected ROIs are visualised in Fig. 10. Since these points provide the most direct visual evidence of atmospheric clarity or obstruction, this method is hypothesized to yield a superior predictive accuracy.

The process follows these steps:

1. Predefined reference locations in the scene, such as distant landmarks or contrast-sensitive objects, are manually selected.
2. The image crops centred on these ROIs are extracted and resized to match the input dimensions of EfficientNetV2M.
3. Feature extraction is performed exclusively on these subregions, removing the broader context of the scene.
4. The extracted features are then processed in the same manner as in the original approach, in combination with tabular meteorological data.

The visualisation of this competitive approach is shown in Fig. 11 and is referred to in the text below as Multimodal (ROIs and Tabular) Model.

While this method is expected to enhance prediction accuracy by focusing on the most informative image regions, it introduces a critical limitation: the approach becomes scene dependent, restricting its applicability to predefined viewpoints with identified reference points. As a result, generalisation to arbitrary surveillance images may be compromised.

3.5.3. Comparative evaluation

To compare these competitive approaches, all models are trained and evaluated using the same dataset and performance metrics. By analysing

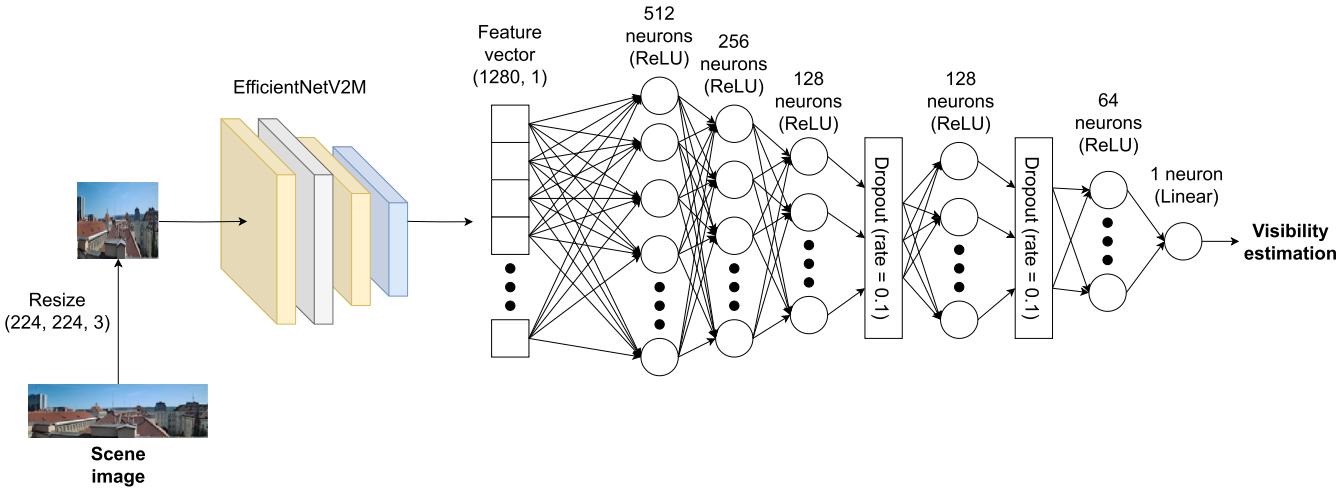


Fig. 8. Architecture of Unimodal (Image) Model used in the ablation study. The EfficientNetV2M-based feature extractor processes the input scene image, generating a high-dimensional feature vector. This vector is subsequently passed through a fully connected network to produce the final visibility estimation. The tabular meteorological data branch is removed in this configuration.

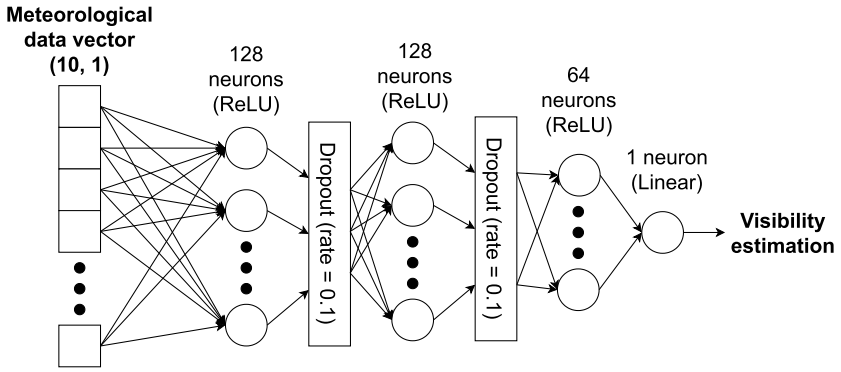


Fig. 9. Architecture of Unimodal (Tabular) Model used in the ablation study. The input consists solely of tabular meteorological data, which are processed through a fully connected neural network. Without the image-based feature extraction, visibility estimation relies entirely on numerical meteorological data.

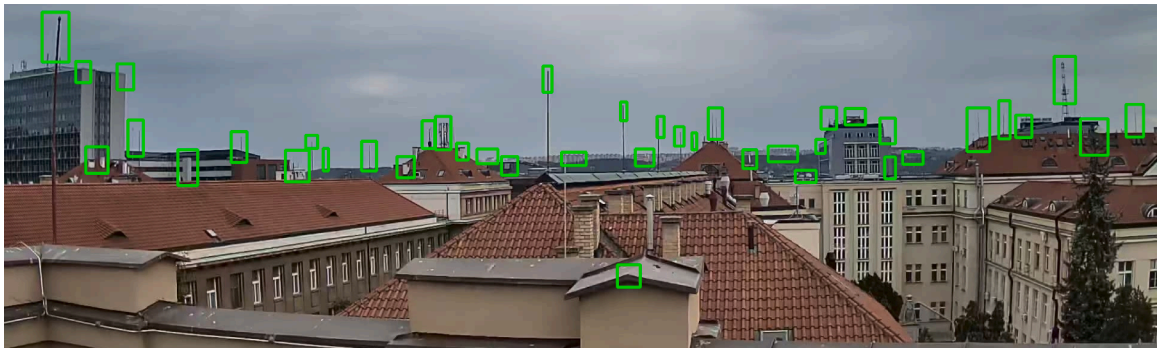


Fig. 10. Visualisation of the regions of interest selected for feature extraction.

their predictive accuracy and generalisation capability, a deeper understanding of the trade-offs associated with multimodal fusion versus specialized feature extraction is obtained.

3.6. Cross-validation

To enhance robust and unbiased performance evaluation of the proposed method with the competitive approaches, a systematic k-fold cross-validation scheme [35] is also implemented. Specifically, the complete dataset of 1000 annotated visibility images, each paired with its corresponding vector of meteorological variables, is divided into $k = 10$ folds. For each fold, the dataset is split into training, validation, and test subsets in the ratio of 70 % : 15 % : 15 %, maintaining the same data proportions across all folds.

To minimize potential bias and ensure broad sample coverage in test subsets, a custom allocation strategy is designed to control test set overlap. While complete non-overlap of test samples across 10 folds would require more data than available, the proposed allocation strategy minimizes test set repetition by tracking sample usage across folds and preferentially selecting samples with the fewest prior appearances in test

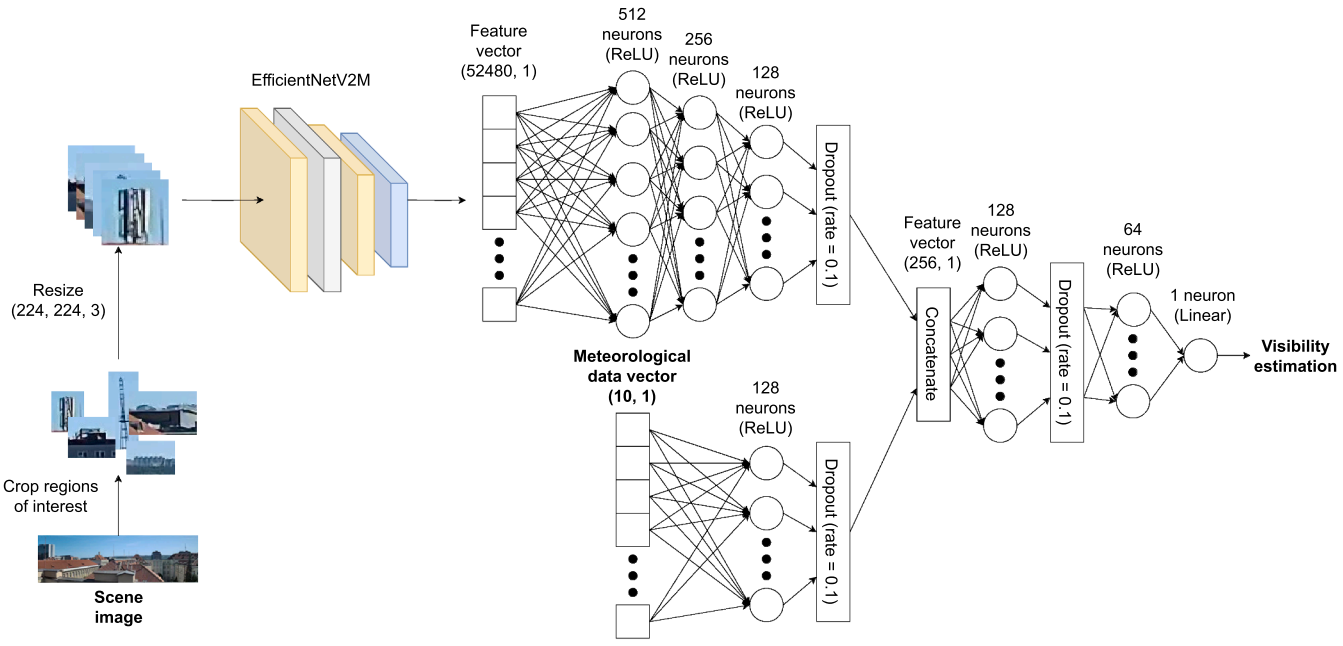


Fig. 11. Architecture of Multimodal (ROIs and Tabular) Model. Instead of processing the entire scene image, predefined regions of interest corresponding to key reference points used for visibility tagging are extracted. Each ROI is individually processed using EfficientNetV2M to generate feature vectors, which are then aggregated and combined with tabular meteorological data.

subsets. This guarantees that each test sample appears in no more than two folds, with the majority being used only once, thus achieving near-optimal coverage of the entire dataset in evaluation.

In each fold, the selected test subset consists of 150 samples, and the remaining data are divided into validation and training subsets to preserve the exact ratio of 15 % and 70 %, respectively. All models evaluated in this study are independently trained and evaluated using the same fold assignments, ensuring fair and consistent comparisons. Furthermore, for each fold, the same training and optimization strategy described in Section 3.3 is applied without modification.

This cross-validation procedure provides a comprehensive performance estimate over the entire dataset and safeguards against overestimating predictive accuracy due to chance partitioning. In total, the results reported in the following section represent the average of ten independent training and evaluation runs.

4. Results and discussion

In this section, the performance of the proposed deep learning model is evaluated for the estimation of atmospheric visibility. The model was trained on a dataset combining image and tabular meteorological inputs, and its predictions were compared to ground truth visibility values using a separate test set. To ensure a comprehensive assessment, all metrics discussed in Section 3.4 were used.

To provide a deeper understanding of the contribution of each input modality, the effect of different architectural variants, and the consequence of tabular meteorological data reduction, an ablation study was conducted with the evaluation over the same metrics. Furthermore, a competitive alternative approach was also evaluated using the regions-of-interest segmentation strategy. The following part presents the numerical results of these experiments in tabular and graphical form, along with a detailed interpretation.

Table 1 summarizes the performance of the proposed model and the competitive approaches described in Sections 3.2 and 3.5. The evaluation is performed on the test set using the quantitative metrics introduced in Section 3.4, covering both absolute and relative performance.

In addition, the same set of metrics was evaluated on the dead-zone-filtered test set, and the results are reported in Table 2. These results are intended to indicate the extent to which the models were influenced by the uncertainty in the ground truth introduced by the gap in the 750–2500 m visibility range.

Each metric provides insight into a different aspect of the behaviour of the model. The size of the model (in kilobytes) reflects the memory footprint, where smaller values are generally preferable due to lower computational and storage demands and the resulting cost-efficiency. The MSE and MAE measure the average deviation of the predictions from the ground truth, with lower values indicating higher accuracy. MSE, compared to MAE, penalizes larger errors more.

The R^2 score quantifies how well the predicted values approximate the true outputs; values closer to 1 indicate stronger explanatory power. Similarly, in the linear regression fit between the predicted and true values, the slope a and the intercept b characterize systematic deviation; optimal performance is reached when a approaches 1 and b approaches 0.

Relative error tolerance metrics assess the percentage of predictions that fall within a given deviation margin from the ground truth. Specifically, the tolerance values ±10 % and ±20 % express the proportion of predictions within 10 % and 20 % of the expected values, respectively. For these metrics, higher percentages are desirable, indicating greater reliability of the model across its prediction range.

In order to improve interpretability and provide a more intuitive understanding of the evaluated behaviour of the proposed model, two additional visualisations are presented.

Fig. 12 shows the prediction error distributions for each of the six models considered. These distributions are calculated as the difference between the predicted and true target values in the test set. For consistency and comparability, the binning range was fixed from −2000 to 2000 m, all histograms being discretized in 60 uniform bins. This visualisation provides a qualitative perspective on the spread and symmetry of the prediction errors and allows to identify potential systematic biases in the model predictions.

The results presented above demonstrate that the proposed multimodal deep learning model achieves strong predictive performance in

Table 1
Comparison of proposed and baseline models based on memory requirements and regression-based evaluation metrics.

Metric	Multimodal (Image and Tabular)	Unimodal (Image)	Unimodal (Tabular)	Multimodal (ROIs and Tabular)	Unimodal (Tabular - reduced)	Multimodal (Image and Tabular - reduced)
Size (kB)	10,179	9,959	165	317,379	156	10,170
MSE (m^2)	129,716	166,607	279,952	74,983	305,430	145,583
MAE (m)	165.4	225.1	278.9	163.1	355.1	197.7
R^2 score	0.8861	0.8537	0.7542	0.9341	0.7318	0.8721
Slope a	0.9160	0.9312	0.8638	0.8952	0.7727	0.8922
Intercept b	243.1	148.5	301.2	203.2	606.9	277.0
±10 %	84.46 %	71.62 %	66.89 %	80.41 %	52.70 %	72.97 %
±20 %	89.19 %	85.14 %	82.43 %	89.86 %	76.35 %	88.51 %

Table 2
Performance of the proposed and competitive models on the dead-zone-filtered test set.

Metric	Multimodal (Image and Tabular)	Unimodal (Image)	Unimodal (Tabular)	Multimodal (ROIs and Tabular)	Unimodal (Tabular - reduced)	Multimodal (Image and Tabular - reduced)
MSE (m^2)	40,422	62,224	126,813	50,894	195,542	44,131
MAE (m)	127.4	181.2	217.4	184.2	299.2	151.7
R^2 score	0.9591	0.9370	0.8717	0.9485	0.8023	0.9553
Slope a	0.9522	0.9889	0.9197	0.8863	0.8348	0.9431
Intercept b	92.7	-68.0	112.2	207.7	395.7	76.6
±10 %	87.23 %	72.34 %	72.34 %	78.72 %	65.96 %	77.66 %
±20 %	94.68 %	90.43 %	88.30 %	91.49 %	84.04 %	92.55 %

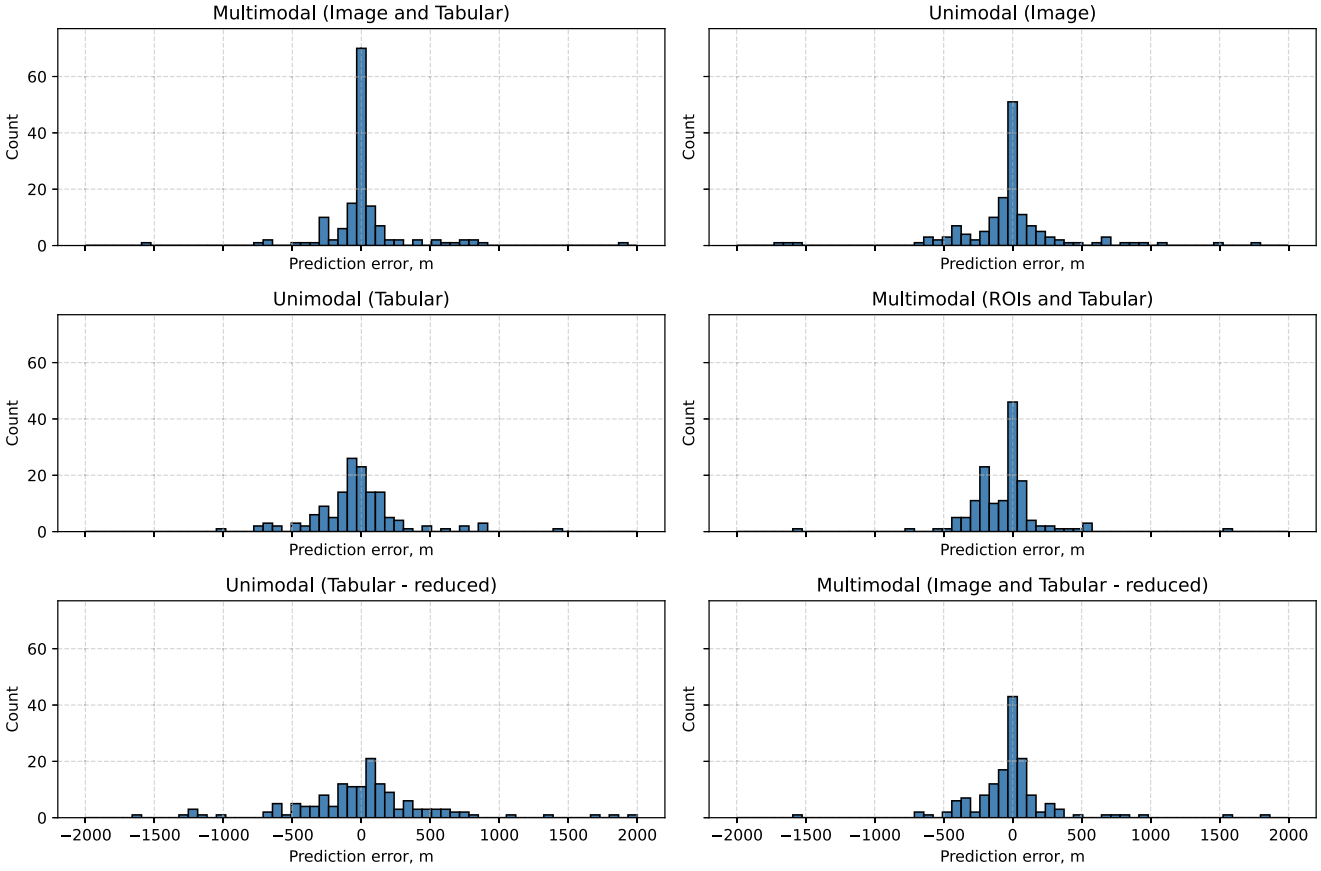


Fig. 12. Prediction error distributions for the six evaluated models.

estimating atmospheric visibility. Across all regression-based metrics evaluated, the model that combines both image and tabular meteorological input data consistently outperforms ablation variants using a single modality. This confirms the initial hypothesis that a joint representation of images and tabular meteorological data provides a more complete and informative context for accurate visibility estimation. Furthermore, the regression slope of 0.9160 and intercept of 243.1 m demonstrate that the model remains well calibrated, with limited systematic bias. Importantly, 84.46 % of test predictions fall within a 10 % relative error margin, and 89.19 % within 20 %, representing the highest performance

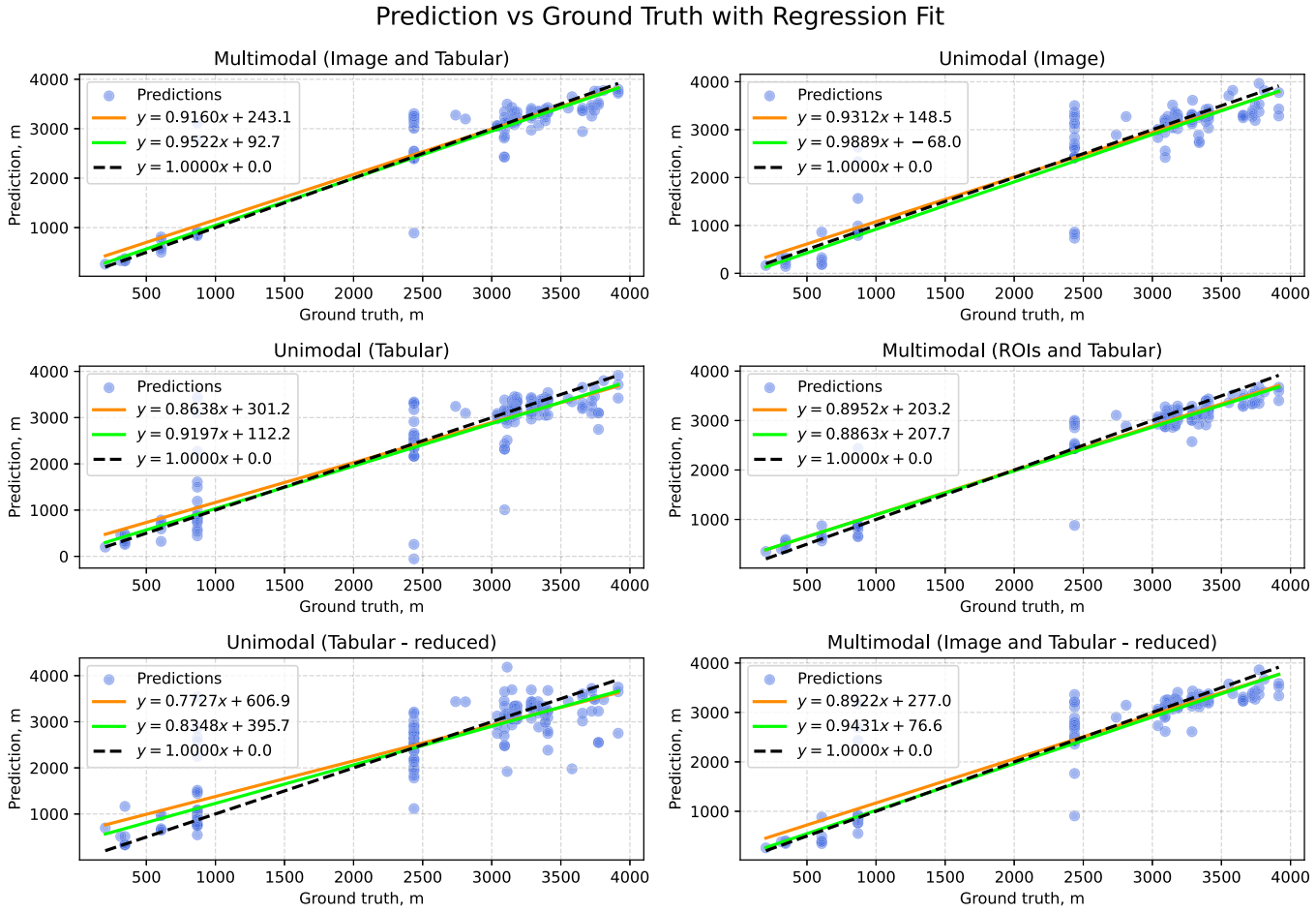


Fig. 13. Scatter plots of predicted versus true values for the six evaluated models..

among all evaluated models in these practically most significant metrics. This confirms the robustness and applicability of the proposed approach in real-world scenarios.

The ablation study provided valuable information on the contribution of each input modality. The Unimodal (Image) Model, while significantly outperforming the Unimodal (Tabular) variant, exhibited higher MSE and MAE error rates than the full multimodal model. However, its regression line ($a = 0.9312$, $b = 148.5$) is closest to the optimal identity line among all the models tested. The Unimodal (Tabular) Model performed substantially worse across all metrics, particularly in its inability to reconstruct scene-specific visibility conditions. This finding underscores the critical role of visual input in accurate visibility estimation, while also highlighting the incremental benefit provided by incorporating meteorological information.

To further examine the role of meteorological input dimensionality, two additional reduced-input variants were evaluated. Both the Unimodal (Tabular - reduced) and Multimodal (Image and Tabular - reduced) models used only four selected meteorological variables: temperature, absolute pressure, atmospheric humidity, and average wind speed. As expected, both reduced models exhibited decreased predictive performance compared to their full-input counterparts. However, the Multimodal (Image and Tabular - reduced) model consistently outperformed the Unimodal (Image) Model across most evaluation metrics, while requiring fewer environmental inputs. This suggests that even a simplified set of meteorological features can improve the accuracy of estimation when combined with visual information. These results emphasize the robustness of the multimodal fusion strategy and support the conclusion that carefully selected tabular features can still provide meaningful gains over visual-only approaches.

Interestingly, Multimodal (ROIs and Tabular) Model delivered the best overall regression performance ($R^2 = 0.9341$; MSE = 74983 m^2; MAE = 163.1 m). These results indicate that, in terms of correlation with ground truth and general trend estimation, Multimodal (ROIs and Tabular) Model provides the most accurate approximation. However, a more detailed examination of the relative error tolerance metrics reveals a less favourable aspect. Specifically, only 80.41 % of the (ROIs and tabular)-based predictions fall within a ± 10 % relative error range, which is noticeably lower than the 84.46 % achieved by the proposed Multimodal (Image and Tabular) Model. This suggests that, although Multimodal (ROIs and Tabular) Model fits the overall trend well, it tends to produce larger local deviations from the actual visibility values more frequently (see also Fig. 12 to support this finding). In addition, the practical limitations of Multimodal (ROIs and Tabular) Model must be emphasized. This model relies on manually selected regions of interest in the image, typically stable reference landmarks used for human annotation, and, as such, is tightly coupled with the specific viewpoint and scene geometry. Implementing this approach at scale would require scene-specific pre-processing and configuration, making it unsuitable for generalised deployment across variable meteorological conditions. In contrast, the proposed model processes the full image automatically and is designed to generalise across scenes without additional annotation effort or geometric knowledge.

To further investigate the influence of ambiguous annotations originating from the dead zone (see Section 2.2), model behaviour on the original (unfiltered) test set (Table 1) and on the dead-zone-filtered test set (Table 2) was performed. This comparison reveals a consistent pattern: after removing samples with annotated visibility in the 750–2500 m interval, most architectures exhibit substantial improve-

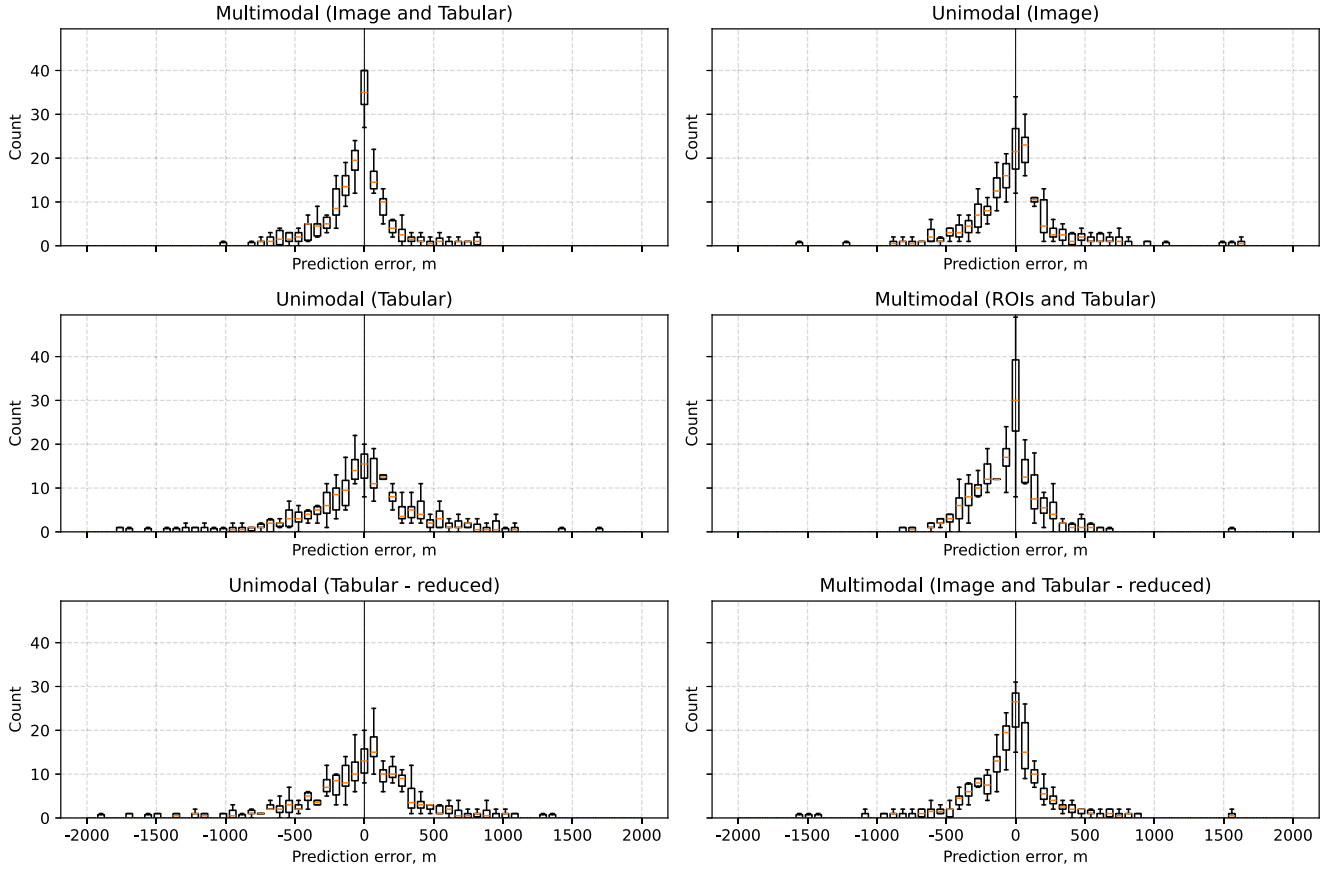


Fig. 14. Box plots of prediction error distributions across ten cross-validation folds.

ments in regression accuracy. These improvements indicate that a substantial portion of large errors on the unfiltered test set can be attributed to samples for which the ground-truth label is uncertain due to the dead zone. By contrast, the Multimodal (ROIs and Tabular) model responds differently to filtering: while its MSE decreases from 74983 m^2 to 50894 m^2, its MAE increases from 163.1 m to 184.2 m and the share of predictions within ±10 % declines from 80.41 % to 78.72 %. Importantly, the ROIs model loses its leading position with respect to the most practically relevant metrics (MAE, MSE and PE_δ) on the dead-zone-filtered partition: the Multimodal (Image and Tabular) architecture attains the lowest MSE (40,422 m^2) and the lowest MAE (127.4 m) on the filtered set, and the highest R^2 (0.9591).

This pattern suggests an interpretation with direct consequences for evaluation practice. Models that depend on manually selected, landmark-centred ROIs encode features that are tightly coupled to the same visual anchors used during annotation. As a result, they are more likely to exploit systematic discretization or annotation artifacts when ground-truth labels are ambiguous (i.e., inside the dead zone). By contrast, the proposed Multimodal (Image and Tabular) model, which operates on the full scene, does not exhibit this increased sensitivity. Its performance improves substantially when gap-affected samples are removed, and it becomes the top performer on the filtered partition. This robustness against annotation-related artifacts indicates a stronger ability to generalise beyond the specific structure of the dataset, suggesting that the proposed model captures more transferable and physically meaningful visibility cues.

For a clear comparison of the summarized findings, Fig. 13 presents scatter plots illustrating the relationship between predicted and actual values across all six models. Each plot includes a regression line fitted on the original test set, a regression line fitted on the dead-zone-filtered test set, and the ideal regression line (slope $a = 1$, intercept $b = 0$) that both fitted lines are expected to approach. These plots help to visually assess the agreement between predictions and targets, and to identify deviations from the ideal identity line, which would correspond to a perfect prediction.

In accordance with the methodology described in Section 3.6, a 10-fold cross-validation was also conducted to further assess the robustness and generalisation capability of the evaluated models. The results of the cross-validation are summarized in Table 3, where each evaluation metric is reported as the average value across all ten folds, accompanied by the standard error of the mean. A graphical overview of the estimation error distributions under the cross-validation scenario is provided in Fig. 14. Unlike Fig. 12, which displays the prediction error distribution based on a single evaluation, Fig. 14 summarizes the variability of prediction errors across the ten cross-validation folds. To construct this visualisation, individual prediction errors were first sorted into 60 uniform bins while the binning range was fixed from −2000 to 2000 m. For each bin, the distribution of counts across folds was captured and visualised using a box plot, thereby reflecting both the central tendency and the variability of error occurrence across the repeated evaluations.

The results confirm the main trends observed in the initial evaluation of the test set. The Multimodal (Image and Tabular) Model continues to demonstrate superior predictive performance, achieving the second lowest cross-validated MSE, the lowest MAE, strong linear correlation, and the highest proportion of predictions within a ±10 % relative error threshold. Although the Multimodal (ROIs and Tabular) Model retains the best MSE and R^2 score, it exhibits a larger spread in error distribution and lower reliability in strict error thresholds.

Reduced-input models again display a decline in performance, particularly the Unimodal (Tabular - reduced) variant, which performs the

Table 3
Cross-validation results (average values accompanied by the standard error of the mean) of the proposed and baseline models.

Metric	Multimodal (Image and Tabular)	Unimodal (Image)	Unimodal (Tabular)	Multimodal (ROIs and Tabular)	Unimodal (Tabular - reduced)	Multimodal (Image and Tabular - reduced)
MSE (m^2)	135,550±10,680	180,604±9,987	270,525±12,437	95,730±15,251	324,363±20,873	167,462±14,980
MAE (m)	207.8±6.6	253.7±9.4	339.0±9.7	208.6±21.0	366.4±9.9	246.2±11.3
R^2 score	0.8798±0.0125	0.8422±0.0110	0.7653±0.0116	0.9210±0.0115	0.7199±0.0153	0.8515±0.0180
Slope a	0.8851±0.0092	0.8598±0.0131	0.8133±0.0183	0.8841±0.0101	0.7962±0.0116	0.8713±0.0073
Intercept b	265.7±29.9	362.9±40.78	497.7±60.2	217.3±23.40	536.9±39.1	305.8±19.3
±10 %	(73.87±1.14)%	(67.75±2.49)%	(55.52±1.10)%	(68.70±4.99)%	(54.57±1.58)%	(68.84±2.12)%
±20 %	(86.87±0.91)%	(82.44±1.20)%	(74.59±1.48)%	(87.55±1.58)%	(73.16±1.07)%	(83.40±1.16)%

worst across all metrics. On the other hand, the Multimodal (Image and Tabular - reduced) Model consistently outperforms the Unimodal (Image) Model in all evaluated criteria, confirming the added value of incorporating even a simplified set of meteorological inputs. Overall, the cross-validation confirms the reliability and generalisation ability of the proposed multimodal fusion strategy under varying data partitions.

5. Conclusion

The evaluation of atmospheric visibility using a combined dataset of images and tabular data, i.e. multimodal input, and its processing with the neural network designed in this work demonstrated that this approach provides accurate and robust visibility estimation in diverse atmospheric conditions. The presented multimodal model showed high consistency with ground truth data, as almost 85 % of visibility evaluations fall within a 10 % relative error margin. These results clearly indicate that combining visual and meteorological modalities leads to significantly improved performance compared to unimodal architectures. Besides, evaluation on the dead-zone-filtered test set confirmed that the proposed multimodal model maintains its leading performance even when samples potentially affected by annotation uncertainty caused by areas without reference points in the observed scene are excluded. This demonstrates that the model does not exhibit increased sensitivity to such uncertainties and thus provides a reliable basis for visibility estimation, i.e. a stronger ability to generalise beyond the specific structure of the dataset. This is a satisfactory result that allows follow-up steps to be taken for the practical application of the outcome of this work.

Several directions for future research emerge from this study. One of the remaining technical challenges is related to the presence of direct light in camera images, such as during sunrise or sunset. This often leads to overexposed visual input and localized prediction errors. However, the multimodal nature of our approach offers a potential mitigation: the visually affected modality can be temporarily down-weighted or excluded, allowing visibility estimation to continue based on the meteorological modality alone, although with reduced accuracy.

Another challenge lies in the significant differences between visibility characteristics during the day and at night. Although our dataset is limited in terms of night-time samples, the proposed multimodal framework provides the conceptual means to extend visibility estimation beyond daylight hours. This aspect will be a key focus of future data collection and validation.

Furthermore, the question of generalisability across different scenes remains open. Since our current models were trained and evaluated on a single location, it is necessary to test whether the proposed approach maintains its performance when deployed in new environments with different terrain, atmospheric behaviour, or lighting conditions.

In addition, further improvements may be achieved by refining the architectures used, particularly the image-processing backbone. While EfficientNet has proven effective, systematic exploration of alternative convolutional or transformer-based encoders, or tailored custom architectures, may lead to gains in performance or robustness.

Another important direction for future work is to address the uncertainty introduced by gaps in the observed scene where no reference points are available. While this study introduced a dead-zone-filtered test set to mitigate their impact during evaluation, a more systematic approach may be needed in the future. Possible strategies include probabilistic or interval-based ground-truth annotation, weighting schemes that explicitly account for annotation uncertainty, or the integration of complementary sensing modalities to reduce reliance on single-scene landmarks.

Finally, a promising direction for future work is to reformulate the task from instantaneous visibility estimation to short-term visibility forecasting. Since the dataset contains temporally ordered image and sensor sequences, the current model could be extended to exploit temporal context and predict future visibility conditions over predefined time horizons. This capability would substantially enhance the practical value of the system in many real-world applications.

In conclusion, the work presented here contributes a practical and scalable solution for automated visibility estimation based on multimodal data fusion. The findings support the continued development of machine learning-based visibility monitoring tools with potential applications in traffic safety, urban infrastructure, or environmental monitoring.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used Writefull Service in order to improve language and readability. After using this service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

CRediT authorship contribution statement

Jitka Kopecká: Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Data curation, Conceptualization; **Dušan Kopecký:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Data curation, Conceptualization; **Dominik Štursa:** Methodology, Investigation; **Zuzana Rácová:** Validation, Investigation, Data curation; **Tomáš Krejčí:** Validation, Software; **Petr Doležel:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Data availability

The data are available at https://zenodo.org/records/15494899.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The work was supported from ERDF “Multi-sector and Interdisciplinary Cooperation in Research and Development of Communication,

Information and Detection Technologies for Control and Signalling Systems (CIDET)" (No. CZ.02.01.01/00/23_021/0008402).

References

[1] H.K. Kim, C.W. Lee, Development of a cost forecasting model for air cargo service delay due to low visibility, Sustainability 11 (16) (2019). https://doi.org/10.3390/su11164390

[2] D. Singh, A.A. AlZubi, M. Kaur, V. Kumar, H.-N. Lee, Deep multi-patch hierarchical network-based visibility restoration model for autonomous vehicles, IEEE Trans. Veh. Technol. (2024) 1–11. https://doi.org/10.1109/TVT.2024.3377605

[3] C. Ekici, I. Teke, Global solar radiation estimation from measurements of visibility and air temperature extremes, Energy Sources Part A 41 (11) (2019) 1344–1359. https://doi.org/10.1080/15567036.2018.1548513

[4] World Meteorological Organization, Guide to Instruments and Methods of Observation (WMO-No. 8), WMO, Geneva, Switzerland, 2018 edition, updated in 2021 edition, Geneva, Switzerland, 2018. WMO-No. 8.

[5] J. Chen, Z. Liu, Z. Yin, X. Liu, X. Li, L. Yin, W. Zheng, Predict the effect of meteorological factors on haze using BP neural network, Urban Clim. 51 (2023). https://doi.org/10.1016/j.uclim.2023.101630

[6] H. Chaabani, F. Kamoun, H. Bargaoui, F. Outay, A.-U.-H. Yasar, A neural network approach to visibility range estimation under foggy weather conditions, Procedia Comput. Sci. 113 (2017) 466–471. https://doi.org/10.1016/j.procs.2017.08.304

[7] W. Wen, L. Li, P.W. Chan, Y.-Y. Liu, M. Wei, Research on the usability of different machine learning methods in visibility forecasting, Atmósfera 37 (2023) 99–111. https://doi.org/10.20937/atm.53053

[8] L.C. Ortega, L.D. Otero, M. Solomon, C.E. Otero, A. Fabregas, Deep learning models for visibility forecasting using climatological data, Int. J. Forecast. 39 (2) (2023) 992–1004. https://doi.org/10.1016/j.ijforecast.2022.03.009

[9] S. Li, H. Fu, W.-L. Lo, Meteorological visibility evaluation on webcam weather image using deep learning features, Int. J. Comput. Theory Eng. 9 (6) (2017) 455–461. https://doi.org/10.7763/IJCTE.2017.V9.1186

[10] D.G. Hadjimitsis, C. Clayton, L. Toulios, Retrieving visibility values using satellite remote sensing data, Phys. Chem. Earth. 35 (1–2) (2010) 121–124. https://doi.org/10.1016/j.pce.2010.03.002

[11] L. Rejfek, K. Juryca, T.N. Nguyen, L. Beran, M. Voznak, Whitening filters application for ionospheric propagation delay extraction, IEEE Trans. Instrum. Meas. 72 (2023). https://doi.org/10.1109/TIM.2023.3279464

[12] Z. Liu, Y. Chen, X. Gu, K.W.J. Yeoh, Q. Zhang, Visibility classification and influencing-factors analysis of airport: a deep learning approach, Atmos. Environ. 278 (2022). https://doi.org/10.1016/j.atmosenv.2022.119085

[13] Z. Gáborčíková, J. Bartok, O.I. Malkin, W. Benešová, L. Ivica, S. Hnilicová, L. Gaál, Artificial intelligence-based detection of light points: an aid for night-time visibility observations, Atmosphere 15 (8) (2024). https://doi.org/10.3390/atmos15080890

[14] H. Wang, K. Shen, P. Yu, Q. Shi, H. Ko, Multimodal deep fusion network for visibility assessment with a small training dataset, IEEE Access 8 (2020) 217057–217067. https://doi.org/10.1109/ACCESS.2020.3031283

[15] R. Babari, N. Hautière, É. Dumont, R. Brémond, N. Paparoditis, et al., A model-driven approach to estimate atmospheric visibility with ordinary cameras, Atmos. Environ. 45 (30) (2011) 5316–5324. https://doi.org/10.1016/j.atmosenv.2011.06.053

[16] F. Carretas, F. Wagner, F.M. Janeiro, Atmospheric visibility and AngstrÖm exponent measurements through digital photography, Meas. J. Int. Meas. Confederation 64 (2015) 147–156. https://doi.org/10.1016/j.measurement.2014.12.041

[17] S. Sabetghadam, F. Ahmadi-Givi, Relationship of extinction coefficient, air pollution, and meteorological parameters in an urban area during 2007 to 2009, Environ. Sci. Pollut. Res. 21 (1) (2014) 538–547. https://doi.org/10.1007/s11356-013-1901-9

[18] Z. Liu, J. Zhang, Y. Yang, Y. Wang, W. Luo, X. Zhou, Enhancing weather forecast accuracy through the integration of WRF and BP neural networks: a novel approach, Earth Space Sci. 11 (10) (2024). https://doi.org/10.1029/2024EA003613

[19] K. Ait Ouadil, S. Idbraim, T. Bouhsine, N. Carla Bouaynaya, H. Alfergani, C. Cliff Johnson, Atmospheric visibility estimation: a review of deep learning approach, Multimed. Tools Appl. 83 (12) (2024) 36261–36286. https://doi.org/10.1007/s11042-023-16855-z

[20] Y. You, C. Lu, W. Wang, C.-K. Tang, Relative CNN-RNN: learning relative atmospheric visibility from images, IEEE Trans. Image Process. 28 (1) (2019) 45–55. https://doi.org/10.1109/TIP.2018.2857219

[21] G. Majewski, B. Szeląg, W. Rogula-Kozłowska, P. Rogula-Kopiec, A. Brandyk, J. Rybak, M. Radziemska, E. Liniauskiene, B. Klik, Machine learning analysis of PM1 impact on visibility with comprehensive sensitivity evaluation of concentration, composition, and meteorological factors, Sci. Rep. 14 (1) (2024). https://doi.org/10.1038/s41598-024-67576-8

[22] D. Sládek, A. Donateo, Application of K-nearest neighbor classification for static webcams visibility observation, Adv. Meteorol. 2023 (2023). https://doi.org/10.1155/2023/6285569

[23] H. Chaabani, N. Werghi, F. Kamoun, B. Taha, F. Outay, A.-U.-H. Yasar, Estimating meteorological visibility range under foggy weather conditions: a deep learning approach, in: 9th International Conference on Emerging Ubiquitous Systems and Pervasive Networks, EUSPN 2018, 141, 2018, pp. 478–483. https://doi.org/10.1016/j.procs.2018.10.139

[24] L. Rejfek, J. Pidanic, D. Stursa, T.N. Nguyen, P.T. Tran, Z. Nemec, T. Zalabsky, Passage detection of a train via a reference point, Lecture Notes Elect. Eng. 1081 (2024) 119–130. https://doi.org/10.1007/978-981-99-8703-0_10

[25] J. Chen, M. Yan, M.R.H. Qureshi, K. Geng, Estimating the visibility in foggy weather based on meteorological and video data: a recurrent neural network approach, IET Signal Proc. (2022). https://doi.org/10.1049/sil2.12164

[26] W.-T. Chu, Y.-H. Liang, K.-C. Ho, Visual weather property prediction by multi-task learning and two-dimensional RNNs, Atmosphere 12 (5) (2021). https://doi.org/10.3390/atmos12050584

[27] R.W. Liu, Y. Lu, Y. Guo, W. Ren, F. Zhu, Y. Lv, AiOENet: all-in-one low-visibility enhancement to improve visual perception for intelligent marine vehicles under severe weather conditions, IEEE Trans. Intell. Veh. 9 (2) (2024) 3811–3826. https://doi.org/10.1109/TIV.2023.3347952

[28] J. Liu, X. Chang, Y. Li, Y. Ji, J. Fu, J. Zhong, STCN-NEt: a novel multi-feature stream fusion visibility estimation approach, IEEE Access 10 (2022) 120329–120342. https://doi.org/10.1109/ACCESS.2022.3218456

[29] M. Song, X. Han, X.F. Liu, Q. Li, Visibility estimation via deep label distribution learning in cloud environment, J. Cloud Comput. Adv. Syst. Appl. 10 (1) (2021). https://doi.org/10.1186/s13677-021-00261-7

[30] X. Wu, D. Hong, J. Chanussot, Convolutional neural networks for multimodal remote sensing data classification, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–10. https://doi.org/10.1109/TGRS.2021.3124913

[31] C. Li, D. Hong, B. Zhang, Y. Li, G. Camps-Valls, X.X. Zhu, J. Chanussot, UrbanSAM: learning invariance-inspired adapters for segment anything models in urban construction, 2025, https://doi.org/10.48550/arXiv.2502.15199

[32] C. Li, B. Zhang, D. Hong, J. Zhou, G. Vivone, S. Li, J. Chanussot, CasFormer: cascaded transformers for fusion-aware computational hyperspectral imaging, Inf. Fusion 108 (2024) 102408. https://doi.org/10.1016/j.inffus.2024.102408

[33] M. Tan, V.L. Quoc, EfficientNetV2: smaller models and faster training, in: 38th International Conference on Machine Learning, ICML 2021, 139, 2021, pp. 10096–10106. https://doi.org/10.48550/arXiv.2104.00298

[34] D. Bäumer, S. Versick, B. Vogel, Determination of the visibility using a digital panorama camera, Atmos. Environ. 42 (11) (2008) 2593–2602. https://doi.org/10.1016/j.atmosenv.2007.06.024

[35] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Springer Series in Statistics, Springer New York Inc., New York, NY, USA, New York, NY, USA, 2001.