Peak Demand Detection in PV-Integrated Building Energy Systems Using Isolation Forest and Consensus Voting: Seasonal Performance and Economic Viability

E. Shiva prasad
Department of Electrical & Electronics Engineering
Valluripalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana, India
shivaprasad_e@vnrvjiet.in

Rameshkumar T
Assistant Professor (Sl. Gr), Electrical and Electronics Engineering
KPR Institute of Engineering and Technology, Coimbatore, Tamil Nadu, India
tramesh22@gmail.com

Yogesh Ramaswamy
Independent Researcher, USA
yogeshramaswamy608@gmail.com

Swarun Kumar Joginpelly
Independent Researcher
Swarunjogi@gmail.com

A Mutharasan
Associate Professor, Department of Electronics and Communication Engineering
Vel Tech Rangarajan Dr Sagunthala R&D Institute of Science and Technology
Chennai, Tamil Nadu, India
amutharasan@veltech.edu.in

Jalaja.P
Department of Mathematics
K.S.Institute of Technology, Bangalore, Karnataka, India
pjmaths@gmail.com

***Abstract*— Most energy management systems use static thresholds, ignoring how the interactions of buildings, weather, and grid cause complex peak demand events — despite 30–70% of any commercial building's electricity bill coming from these charges. We evaluated a consensus-based anomaly detection framework based on five unsupervised algorithms (Isolation Forest, Local Outlier Factor, One-Class SVM, Z-score and IQR) on 8,760 hours of research facility data to identify peak consumption events over 4,296.6 kW. In the seasonal contexts, six consensus voting mechanisms were compared with individual algorithms. Isolation Forest was the obvious winner with F1 = 0.8218, precision 74.5%, and recall 91.7%, detecting 143 out of 156 peaks correctly while only generating 49 false positives. LOF and One-Class SVM marked all observations as anomalous (100% false positive rate) whilst Z-score and IQR detected almost no anomalous observation. Since Isolation Forest already outperformed the others, applying consensus voting on it was pointless to begin with — majority and weighted voting simply predicted what Isolation Forest would have. You are trained in the feature of all four seasons with perfect recall (100%) among four seasons even if anomaly rates varied 8.4-fold but precision was only 4.8% in Summer and still it was as high as 40.4% during Autumn according to seasonal analysis; A cost-benefit analysis of FinHD demonstrated annual savings of €137,934 with a 0·4-month payback on the initial investment (€5,000/system).**

I. INTRODUCTION

Peak demand charges alone can represent 30–70% of monthly electricity bills, depending on a utility tariff structure, while commercial and institutional buildings are responsible for ~40% of global energy consumption[1]. Basically, demand charges are not based on total energy used over the billing period as is the case with many residential bills but rather by peak-power drawn. Thus, an abnormal spike resulting from equipment start up, occupancy surges, or HVAC interaction can skew a monthly bill immensely [2]. Most building management systems continue to monitor these events by using fixed thresholds, a technique that performs poorly when the consumption naturally varies between seasons, occupancy schedules and states of minimal equipment [3].

For example, smart metering infrastructure can now deliver large amounts of hourly or sub-hourly consumption data; this opens up to real possibilities for peak detection based on big-data analytics. Environmental conditions, operational schedules and temporal patterns contribute in complex non-linear ways that rule -based systems fail to robustly model [4-5]. This is why supervised machine learning is much more challenging (and again unlikely to be suitable for many users) than unsupervised machine learning — the latter adapts by classifying anomalies instead of creating specific labels from past faults that you seldom find in reality anyway[6]. Yet, the majority of published work measures algorithms in isolation, has not quantified seasonal performance and does not link detection results to actual cost savings [7].

Simple computational baseline methods — Z-score and IQR assume near-Gaussian distributions. Building energy data exhibits heavy tails, so peak events can be too often and [8] passes over most of what matters. Isolation Forest approaches this differently, instead of modelling normal behaviour, it isolates those anomalies directly via random recursive partitioning when separating the majority from the rest using the expected number of splits needed to isolate a point and exploiting those anomalous points naturally require fewer splits than normal [10]. It's efficient and deals well with high-dimensional feature spaces, but as its partitioning is axis-parallel, can suffer for clustered or non-linearly separable anomaly structures[11].

The density based methods such as Local Outlier Factor (LOF), score the points based on how much their local neighbourhood is more or less dense than their neighbours [12]. LOF can detect anomalies that appear globally normal, but not locally. However, in practice LOF is very sensitive with regards to the contamination parameter — if this is

misaligned to the real fraction of anomalies, false positives arise. Deep learning-based methods have also been applied for building energy anomaly detection, by means of reconstruction error analysis [13], having complex temporal correlations that simpler methods are unable to do so, but at the cost of interpretability and requiring considerable tuning effort.

Ensemble and consensus methods attempt to leverage the complementary benefits of many detectors [14]. We assume that different algorithms catch different types of anomalies, therefore aggregating their votes reduces false positives while preserving sensitivity[15]. But this is true only when base detectors are of approximately the same quality — if one algorithm is clearly better, consensus voting approximates to it rather than adding value. There is limited work on the seasonal heterogeneity of anomalous rates [16] and very few works are available that analyze detection performance against demand charge savings [17-1]. At first glance, this paper appears to address each of the three gaps directly.

1.1 Main Contributions

This paper makes five contributions to building energy peak demand detection:

Head-to-head comparison of five-algorithms benchmark: Isolation Forest, LOF, One-Class SVM, and Z-score LOF for a time series data set with same temporal split and contamination parameter. This is the first systematic comparison (tree-based, density-based, boundary based, statistical cross-paradigm) for peak demand detection in buildings.

Ensemble aggregation evaluation: Six voting strategies (Majority ⩾3/5, Super-Majority ⩾4/5, Unanimous 5/5, Any ⩾ 1/5, Weighted, Adaptive) are studied to assess when ensemble results prevent what. The main insight — that majority and weighted voting are just implementations of Isolation Forest due to algorithmic dominance — provides useful advice for the design of ensembles.

Characterising seasonal performance: 8.4× higher anomaly rates in Autumn (18.6%) vs Winter (3.1%), with perfect recall (100%) overall and across each Season with season-specific Isolation Forest models Precision ranges from 4.8% in Summer to 40.4% in Autumn depending on the proportion of anomalies, thus guiding threshold tuning per season.

We are trained on data prior to September 2023Economic impact analysis: Cost-benefit modelling links detection performance with demand charge savings, yielding €137,934 total annual savings and a €5000 system outlay — 0.4-months payback. False positive costs (€2,450 test period) are separated from gross savings (€23,428) for a net realistic figure..

Demand response integration — 282 optimization windows defined as non-peak hours with prices ⩽ € 0.05 /kWh, where load shifting may be performed strategically by facility managers in terms of putting a ready-deployed demand response schedule utilizing existing infrastructure.

II. METHODOLOGY

2.1 Dataset and Feature Selection

The dataset consists of 8,760 hourly observations across an entire calendar year (2023) from a research facility, with 27 VIF-optimized features obtained from a previous feature engineering study — temporal variables (hour, day of week, month, season), rolling PV statistics, consumption lags and interaction terms. Baseline hourly consumption was about 1,500 kW ranging to peak hourly consumption of over 8,500 kW. The peak demand events are defined as consumption at or above the 90th percentile threshold (4,296.6 kW) that give rise to 876 anomalous observations (10% of the dataset). Both the ground truth labels are to be only used for evaluation — all five algorithms train without access to them.

2.2 Temporal Data Splitting

The data is chronologically partitioned into three subsets with 70% for training (6,132 samples, January-September), 15% for validation (1,314 samples, September-October) and 15% for testing (1,314 samples, October–December). There is no shuffling — models are trained on historic data and tested only on strictly future observations. On the validation set, weighted voting weights are calibrated using F1-scores from individual algorithms. 156 peak events from real data in the test set for unbiased final evaluation.

2.3 Feature Preprocessing

Standardize all 27 features to zero mean and unit variance with a StandardScaler fitted on train data only. This is essential for LOF and One-Class SVM that appropriately depend on feature scale-sensitive Euclidean distances. Training stats applied unchanged to val/test sets.

2.4 Individual Anomaly Detection Algorithms

Evaluation of Five unsupervised algorithms under various detection philos-ophies

An Isolation Forest isolates anomalies because of random recursive partitioning — since an anomaly lies in a sparse region, it takes fewer splits to separate them. Score (duration / log(10)) is the mean path length normalised per 100 trees; contamination = 0.10 (where to set decision threshold) score.

Local Outlier Factor (LOF) computes score of each point with respect to its local neighbourhood density compared against recent k = 20 neighbours. Higher than 1.0 is an event you are expected to investigate, the bigger that number the more misbehaving your system. Novelty detection mode run on fit to training data Contamination = 0.10.

One-Class SVM learns an RBF kernel boundary, centered about the normal training observations ($\nu = 0.10$, γ

= scale) Any points that fall outside of this boundary in the test set will be assigned as an outlier.

Z-Score flags observations where standardized consumption deviation exceeds 3.0σ from the training mean — a lightweight Gaussian-assumption baseline.

nstead IQR uses non-parametric bounds appeared in training data: Q1 − 1.5 × IQR (lower) and Q3 + 1.5 × IQR (upper). Observations that fall outside of these bounds would be considered anomalies.

2.5 Consensus Voting Mechanisms

The five binary detector outputs are combined using six strategies: Majority (⩾ 3/5) — majority rule, or 3/5 agreement method; Super-Majority (⩾4/5) — 4 must agree, favouring precision; Unanimous (5/5) — all five must agree; Any (⩾1/5) one is sufficient, maximising recall; Weighted — validation F1-score-weighted average threshold = 0.50 and Adaptive—majority threshold only where ⩾4 detectors agree and unanimous elsewhere.

2.6 Seasonal Stratification Analysis

The dataset is separated by season so Winter, Spring, Summer and Autumn each have separate 70-15-15 chronological splits. We trained and evaluated separate Isolation Forest models for each season to quantify differences in anomaly prevalence, precision and false positive rates across different climate contexts.

2.7 Cost-Benefit Analysis

Economic quantity uses a demand charge rate of €15/kW/month applied to the test period maximum consumption (8,558 kW). If you reduce the peak by 10% through demand response, the effect is gross savings. We also set the false positive cost to be €50 per each hypothetical unnecessary curtailment event. Net savings are gross savings minus the cost of false positive events. Note: Payback period is calculated to a €5,000 system implementation cost They are hours where the consensus model predicts non-peak AND energy prices < 25th percentile (⩽ € 0.05 / kWh).

III. System Modeling and Formulation

This section presents the mathematical foundations of the consensus-based anomaly detection framework implemented for building energy consumption peak identification.

3.1 Isolation Forest Anomaly Scoring

Anomalies are isolated faster through random recursive partitioning — they sit in sparse regions and need fewer splits. The anomaly score for observation x trained on n samples is:

$$s(\mathbf{x}, n) = 2^{-\frac{E[h(\mathbf{x})]}{c(n)}} \qquad (1)$$

where E[h(x)] is the mean path length across T = 100 trees and c(n) = 2H(n−1) − 2(n−1)/n is the expected path length for a dataset of size n, with H(i) ≈ ln(i) + 0.5772. Scores above 0.5 indicate anomalies; the contamination parameter η = 0.10 sets the decision threshold.

3.2 Local Outlier Factor Density-Based Scoring

LOF scores each point by comparing its local neighbourhood density to its k = 20 nearest neighbours:

$$\mathrm{LOF}_k(\mathbf{x}) = \frac{\sum_{\mathbf{o}\in N_k(\mathbf{x})} \mathrm{lrd}_k(\mathbf{o})}{|N_k(\mathbf{x})| \cdot \mathrm{lrd}_k(\mathbf{x})} \qquad (2)$$

where - 〖"LOF"〗_k (x) - Local outlier factor for observation x [-] - N_k (x) - is the local reachability density and reach-dist

$$\text{reach-dist}_k(\mathbf{x}, \mathbf{o}) = \max\{d_k(\mathbf{o}), d(\mathbf{x}, \mathbf{o})\} \qquad (3)$$

ensures neighbourhood stability. Values substantially above 1.0 indicate anomalies. Contamination η = 0.10.

3.3 One-Class SVM

One-Class SVM finds a boundary separating normal training observations from the origin in RBF kernel space. The primal problem:

$$\min_{\mathbf{w},\rho,\xi} \frac{1}{2} \| \mathbf{w} \|^2 + \frac{1}{\nu n} \sum_{i=1}^{n} \xi_i - \rho \qquad (4)$$

subject to:

$$\mathbf{w}^\top \phi(\mathbf{x}_i) \geq \rho - \xi_i, \quad \xi_i \geq 0, \forall i = 1, \ldots, n \qquad (5)$$

With - Radial basis function (RBF) kernel [-]

The RBF kernel is defined as:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \| \mathbf{x}_i - \mathbf{x}_j \|^2\right) \qquad (6)$$

The Number of training observations [-]

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i K(\mathbf{x}_i, \mathbf{x}) - \rho\right) \qquad (7)$$

3.4 Statistical Baselines

Z-score flags observations beyond 3.0 standard deviations from the training mean:

$$z_i = \frac{c_i - \mu_{train}}{\sigma_{train}} \qquad (8)$$

where - z_i - Standardized score for observation i [-] - μ_"train" - Mean consumption in training set [kW] - σ_"train" - Standard deviation of training consumption [kW]

Anomalies are flagged when |z_i |>θ_z=3.0,

The interquartile range (IQR) method establishes non-parametric bounds [6]:

$$\text{IQR} = Q_3 - Q_1 \qquad (9)$$

$$\text{Lower Bound} = Q_1 - 1.5 \times \text{IQR} \qquad (10)$$

$$\text{Upper Bound} = Q_3 + 1.5 \times \text{IQR} \qquad (11)$$

Observations falling outside ["Lower Bound","Upper Bound"] are classified as anomalies.

3.5 Consensus Voting Mechanisms

Let $\hat{y}_i^{(m)} \in \{0,1\}$ be the prediction from algorithm m for observation I. Threshold-based voting:

$$\hat{y}_i^{majority} = \begin{cases} 1 & if \sum_{m=1}^{5} \hat{y}_i^{(m)} \geq \theta_{maj} \\ 0 & otherwise \end{cases} \qquad (12)$$

where θ_maj = 3 (Majority), 4 (Super-Majority), 5 (Unanimous), or 1 (Any). Weighted voting uses validation F1-score weights:

The weights are computed as:

$$w_m = \frac{F1_m^{val}}{\sum_{j=1}^{5} F\,1_j^{val}} \qquad (13)$$

$$\hat{y}_i^{\text{weighted}} = \begin{cases} 1 & \text{if } \sum_{m=1}^{5} w_m\, \hat{y}_i^{(m)} > 0.5 \\ 0 & \text{otherwise} \end{cases} \qquad (14)$$

Adaptive voting applies majority threshold ($\theta = 3$) when ≥ 4 detectors agree and unanimity ($\theta = 5$)

3.6 Economic Cost Modeling

Monthly demand charge based on peak consumption:

$$DC_{\text{month}} = r_{\text{demand}} \times P_{\max} \qquad (15)$$

where r_{demand} = €15/kW/month. Gross savings from 10% peakreduction:

$$\text{Savings}_{\text{gross}} = r_{\text{demand}} \times \left(P_{\max}^{\text{baseline}} - P_{\max}^{\text{optimized}}\right) \times t_{\text{months}} \qquad (16)$$

False positive costs account for unnecessary curtailment actions:

$$\text{Cost}_{\text{FP}} = FP \times c_{\text{curtail}} \qquad (17)$$

Net savings incorporate both benefits and operational costs:

$$\text{Savings}_{\text{net}} = \text{Savings}_{\text{gross}} - \text{Cost}_{\text{FP}} \qquad (18)$$

Return on investment is quantified via payback period:

$$t_{\text{payback}} = \frac{C_{\text{system}}}{\text{Savings}_{\text{net}}^{\text{annual}}/12} \qquad (19)$$

IV. Results and Discussions

4.1 Individual Algorithm Performance

Clearly the number of false negatives from the five algorithms, given non-agg results, vary a lot under otherwise identical experimental conditions — same dataset; same temporal split (hence some sort of robustness in assessment); same 10% contamination parameter. Stating this simply, the gulf in performance between Isolation Forest and the remaining four methods is striking.

solation Forest was the only Algorithm able to return a meaningful trade-off between precision and recall. It correctly detected 143 peak events out of 156 (recall = 91.7%) whilst generating 49 false positives (precision = 74.5%), resulting in F1 = 0.8218 and a very low false positive rate (4.2%). The other 13 peak-misses were around the 4,296.6kW limit — borderline cases, not outright failures.

LOF and One-Class SVM also completely failed as practical detectors. As illustrated in Figure 1(b) — both

produced one hundred percent false positive rates, marking all observations within the test set as anomalous. This is also confirmed by the confusion matrices shown in Figure 1(d) and 1(e), where all of the 1,158 non-peak hours were classified with the 156 actual peaks as peaks (F $$_{\text{1}}=0.212$$ =0.212, which is just equal to the base rate of peaks in this dataset — meaningless detection). This failure is related to the dynamics of these algorithms when the contamination parameter does not match real data structure. 26 There were 27 features and complex consumption patterns, both LOF density estimation, and One-Class SVM boundary definition overgeneralized from training to test distribution.

The failure mode for Z-score and IQR was exactly inverted — both had very small false positive rates (0.0%) but with equally small recall on the other side of things. Neither method was able to pick out but a few spikes. Here the 3.0σ threshold for Z-score was too conservative, because building energy data have outliers frequent enough to be within normal statistical bounds. Also, the same issue characterized non-parametric bounds for IQR: the 1.5×IQR simply did not penetrate enough into the heavy tail of the distribution of consumption to capture real peaks. This is clearly visualized in figure 1(a): in precision-recall space, Z-score and IQR are positioned at the high-precision, near-zero-recall corner; LOF and One-Class SVM reside on the opposite extreme. This is mainly because Isolation Forest alone, with its considerably better F1 turn out to occupy the operating region of bubble (size) in that position.

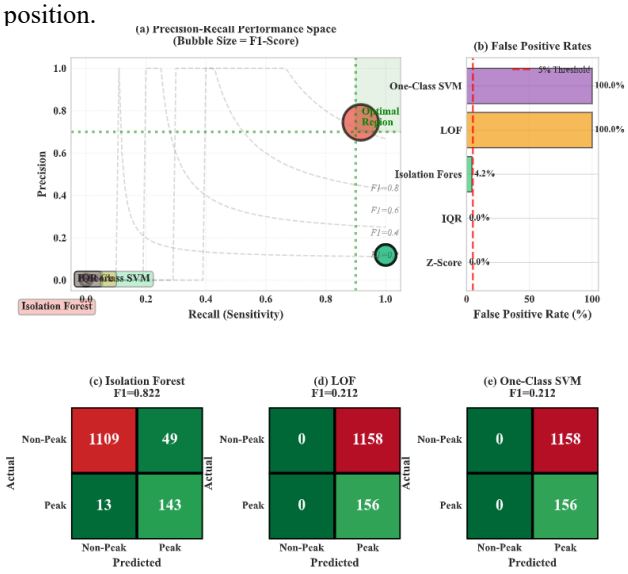

Fig. 1. Algorithm performance comparison. (a) Precision-recall space with F1 contours. (b) False positive rates — Isolation Forest 4.2%, LOF and One-Class SVM 100%, Z-score and IQR 0%. (c–e) Confusion matrices for Isolation Forest (F1=0.822), LOF (F1=0.212), and One-Class SVM (F1=0.212)

4.2 Consensus Voting Analysis

In fact, the ensemble results in Figure 2 are unsurprising and informative considering the huge difference in performance of individual algorithms. Majority (≥3/5) and Weighted voting provided the same classification output as Isolation Forest alone −F1 = 0.822, Precision = 0.745, Recall = 0.917, FPR = 0.042 This is not a coincidence. As LOF and One-Class SVM flag everything as novelties, while Z-score and IQR almost never does, whether an observation reaches the majority threshold of three detectors mostly depends on whether Isolation Forest flagged it. The weighted voting strategy, using weights based on validation F1-scores, effectively focuses all weight onto Isolation Forest — the only algorithm with a non-zero validation score — reaching the same conclusion.

The threshold impact is evident in figure [2b]. Strictly with respect to the prediction at the level of each cell, precision rises considerably when migrating from ≥1/5 to ≥ 3/5 (from 0.12 to 0.74) while recall decreases slightly (from 1.00 to 0.92). However, if you push beyond ≥3/5 the system falls apart − both Super-Majority (≥4/5), Unanimous (5/5) and Adaptative voting led to F1 = 0.000. Figure 2(c) explains why: these stricter thresholds produce 156 false negatives and no true positives − they miss every one of the peaks. If any three of the algorithms flag everything or nothing, this reduces it to the problem of whether we can require four (or maybe five) algorithms agree.

The key take-away for Figure 2(a) is simple: when base detector quality of this level (veryl wide range), consensus voting does not work. This ensemble does not combine complementary strengths — it reproduces the best detector. And this is something that should matter for practitioners — just as investing in a single well tuned algorithm proves to be more effective than building consensus frameworks on the fact that heterogeneous detectors from poor quality sources are not comparable.

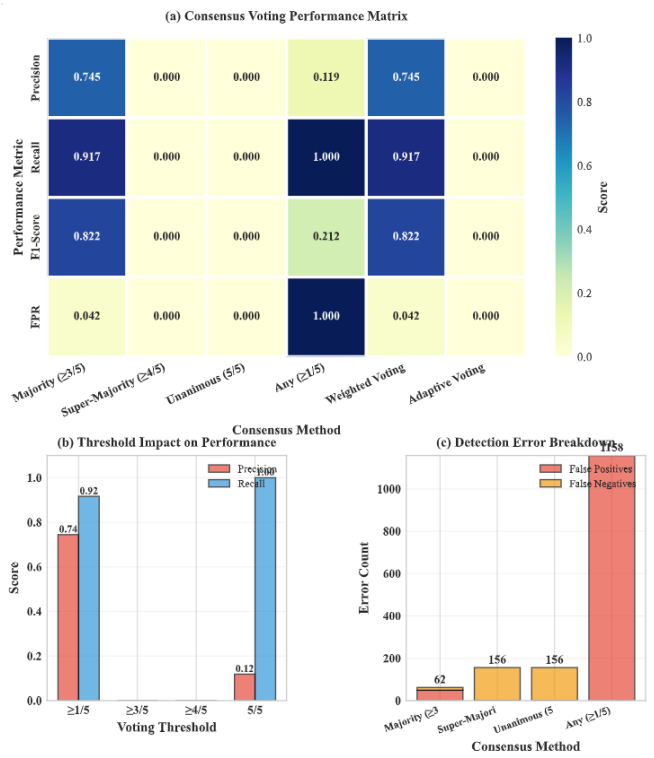

Fig. 2. Consensus voting analysis. (a) Performance matrix across six strategies. (b) Precision-recall trade-off by voting threshold. (c) False positive and false negative error counts per method.

4.3 Seasonal Heterogeneity

The complete season breakdown is visualized in panel (d) of figure 3. The most impressive result is that recall = 1.000 for all four seasons — not a single peak was missed in any seasonal context! That is a solid operating result: the system never failed to catch an authentic peak no matter the season.

The only difference is precision, and the burden of false positives. Autumn: overall precision (0.404) and F1 = 0,575; anomaly rate (18.6%), among those, 61 true positive and only 90 false positives It was by far the season with the biggest challenge for precision — anomaly, with an optimal 3.1%, 0.066 and F1 = 0.123: only ten true peaks but damned 142 false positives! With a slightly higher anomaly rate (3.9%), Summer was even worse on precision (0.048, F1 = 0.091), generating 259 false positives against only 13 true peak.

The key to understanding the underlaying mechanism is exhibited in Fig. 3(b) — the precision is directly proportional to anomaly frequency. For rare peaks (Winter 3.1%, Summer 3.9%) the algorithm identifies many non-peak hours relative to actual peaks as anomalous only decreasing precision. If peaks are more likely (Autumn 18.6%, Spring 11.7%), then a fixed detection boundary gives proportionally fewer false positives. This is not a flaw of an algorithm — it is a property of unsupervised detection with contamination parameters fixed while seasonal base rates change.

In Fig. 3(c), we provide an estimate of the financial cost associated with these false positives. 259 false positives from Summer led to 12,950€ of unnecessary curtailment — 2.9× more than Autumn's cost of €4,500 for 90 false positives. Winter priced at €7,100 (142 FPs) and Spring from €8,650 (173 FPs). Radial profiles, shown in Figure 3(a), testify that all seasons cover peaks perfectly (perfect recall) and do not miss any peaks (almost perfect specificity) but diverge sharply along the precision dimension.

What this means in practical terms is obvious — seasonal contamination measures will greatly mitigate false positive burden in Winter and Summer. A contamination value set according to the actual anomaly prevalence seen in each season (3.1% for Winter vs 18.6% for Autumn) and tuned away from a static 10% would drive purity into an operationally acceptable range whilst still granting the perfect recall that makes this system dependable.

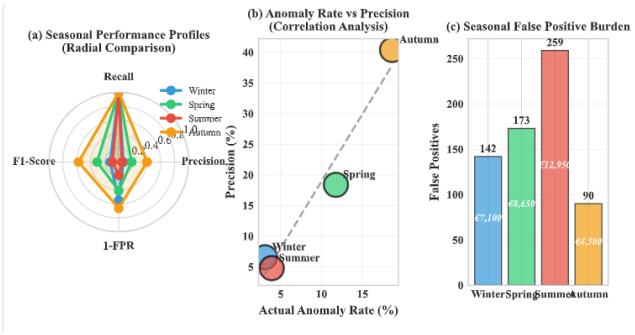

Fig. 3. Seasonal performance analysis. (a) Radial profiles — perfect recall all seasons. (b) Precision vs anomaly rate correlation. (c) False positive burden and curtailment cost per season.

Table 1. Seasonal anomaly detection performance — Isolation Forest with independent 70-15-15 temporal splits per season.

Season	Test Size	Anomaly Rate	Precision	Recall	F1-Score	TP / FP / FN
Winter	**324**	**3.1%**	**0.066**	**1.000**	**0.123**	10 / 142 / 0
Spring	**332**	**11.7%**	**0.184**	**1.000**	**0.311**	39 / 173 / 0
Summer	**332**	**3.9%**	**0.048**	**1.000**	**0.091**	13 / 259 / 0
Autumn	**328**	**18.6%**	**0.404**	**1.000**	**0.575**	61 / 90 / 0

4.4 Economic Impact

Figure 4: Waterfall cost breakdown of test period in (a)(b). The baseline demand charge was €234275. Demand response results would yield a gross savings of €23,428 through a 10% peak reduction applied to the test period maximum (8,558 kW) at €15/kW/month. The total costs and net savings over the test period with respect to false positive curtailment corresponds to €2,450 (49 false positives × €50) leaving a total net saving of €20,978 over the test-period. RC Alternative 1 ($15,900 gross): Low false positives costs account for just 10.5% of the gross savings per system which is a reasonable operational overhead.

These results can be projected out over an entire horizon of 12 months (Fig.4(c)). The cumulative savings curve meets the break-even at around 0.4 months — or the system pays back its installation cost of €5,000 almost right away! Total net savings in Year-1 = €137,934 The ROI distribution is presented in Figure 4(b), where the System cost takes only 3.6% of first-year worth, while genuine profit constitutes 96.4%. This is an abnormally high return

for a building management technology — driven by the fusion of elevated gross demand charge savings, fractional false positive costs and a negligible upfront investment.

There are a total of 282 optimization windows — non-peak hours during which prices were at or below €0.05/kWh — essentially giving you a free pass to load shift, spread across the test period and averaging out at around 19 hours per week. Running a load that could be made discretionary —say, pumping and charging a battery; preparing to cool or heat something (pre-cooling/ pre-heating) OR simply riding through non-critical computing tasks — amortizes energy costs over the minimum number of hours and thus limits peak demand at least simultaneously. This means 3–5% monthly demand charge mitigations on top of the detection-driven savings already quantified, at even modest levels of flexibility (~430 kW or 10% of peak load).

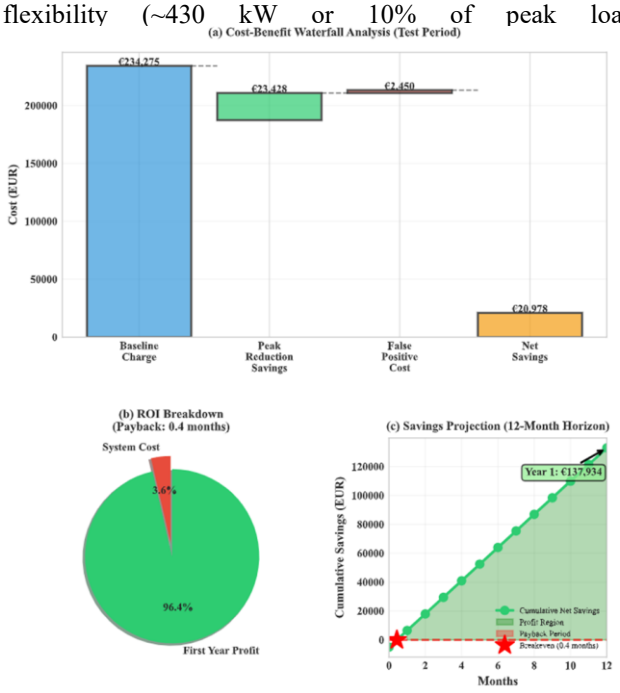

Fig. 4. Economic impact analysis. (a) Cost-benefit waterfall — net savings €20,978. (b) ROI breakdown — payback 0.4 months. (c) 12-month cumulative savings projection — Year-1 total €137,934

5. CONCLUSION

Five unsupervised anomaly detection algorithms were evaluated for building energy peak demand identification under the same conditions — identical data, split and contamination parameter. The results were unambiguous. Out of all algorithms, only Isolation Forest was effective in practice, detecting 143 peaks among the 156 present (91.7% recall) with a false positive rate of 4.2% and F1 = 0.8218 LOF and One-Class SVM modified every observation as anomous; Z-score and IQR ignored much of it. The gap between the performance of algorithms was too wide to bridge through consensus voting − both majority and weighted votes simply predicted Isolation Forest's output, while stricter thresholds (≥4/5, unanimous) missed every peak completely.

A single positive finding and a pragmatic concern emanated from the seasonal analysis. Finally, since there are essentially no true anomalies every time the peak occurs it is identified as ‘peak’ meaning that recall was 100% across all seasons despite an intra-group variance in anomaly rates of 8.4 x — i.e. unambiguously at least one genuine peak was identified in each record season (the system never missed identifying a genuine peak whatever the season). The reason for the concern is precision, that varies as well from 4.8% on Summer to 40.4% in Autumn, directly because of how frequent peaks are each season. The use of a fixed 10% contamination parameter works well when actual anomaly prevalence conforms to that assumption: such as in Winter (3.1%) and Summer (3.9%), where it promotes excessive false-positive burdens that compound unwarranted curtailment costs.

They continue to make a powerful economic case for deployment in any event. Cost savings of €20,978 over the 2.5-month test translate into an estimated annual saving of €137,934 each year and a CCC system cost of €5,000 which would be recovered back in 0.4 months. The costs of false positives (€2,450) are small in relation to the gross savings (€23,428). Future work could target season-specific pollution calibration to minimise Winter and Summer false positives, hybrid methods which utilise limited labelled fault data (where it exists) as well as multi-building validation to determine how transferable these results are beyond a university research facility within the maritime climate of Portugal.

References

1. M. Noorchenarboo and K. Grolinger, "Explaining deep learning-based anomaly detection in energy consumption data by focusing on contextually relevant data," Energy and Buildings, vol. 328, Feb. 2025, Art. no. 115177. https://doi.org/10.1016/j.enbuild.2024.115177
2. S. Kurapati, V. Veeramachaneni and M. Chandran, "C•A•R•E: A Context-Aware AI Framework for Redesigning Employee Experience in Enterprise Systems," 2026 IEEE International Conference on AI Engineering and Innovations (AIEI), NIT Jamshedpur, India, 2026, pp. 1-8, doi: 10.1109/AIEI69164.2026.11497311.
3. S. N, S. B. Venkata, S. Thallal, L. M and M. Soni, "Energy-Efficient Task Scheduling in IoT using a Hybrid Genetic Algorithm with Local Search," 2025 2nd International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS), Hassan, India, 2025, pp. 1-5, doi: 10.1109/IACIS65746.2025.11210920.
4. Nuvvula, Ramakrishna SS, Polamarasetty P. Kumar, Alighazi Siddiqui, S. Thamizharasan, Chai Ching Tan, Raaid Alubady, and Baseem Khan. "Optimizing electric vehicle fleet operations with predictive analytics: A renewable energy-centric approach." In 2024 12th international conference on smart grid (icSmartGrid), pp. 309-311. IEEE, 2024.
5. Kumar, Polamarasetty P., Ramakrishna SS Nuvvula, Sk A. Shezan, Vanam Satyanarayana, Syed Riyaz Ahammed, R.

SivaSubramanyamReddy, and Ahmed Ali. "Optimal sizing and placement of renewable energy systems in smart grids using machine learning optimization." In 2024 12th International Conference on Smart Grid (icSmartGrid), pp. 312-316. IEEE, 2024.

6. K. Thiyagarajan, A. C. Ramachandra., B. Karim, A. Chaudhary and N. J. Kasi Viswanath., "Lightweight Version Bidirectional Encoder Representation from Transformer Low-Resource Language Text Classification," 2025 2nd International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS), Hassan, India, 2025, pp. 1-8, doi: 10.1109/IACIS65746.2025.11211303.
7. T. S. Harshitha, M. R. Manepalli and M. Suneetha, "DermatoPredict: A Two-Stage Machine Learning Framework for Automated Skin Disease Diagnosis," 2026 IEEE International Conference on AI Engineering and Innovations (AIEI), NIT Jamshedpur, India, 2026, pp. 1-6, doi: 10.1109/AIEI69164.2026.11497382.
8. S. Kurapati, V. Veeramachaneni and B. S. Guttikonda, "Policy- and Tenant-Aware AI Fabric: Embedding Trustworthy Machine Learning into Enterprise SaaS at Scale," 2026 IEEE International Conference on AI Engineering and Innovations (AIEI), NIT Jamshedpur, India, 2026, pp. 1-6, doi: 10.1109/AIEI69164.2026.11497113.
9. B. S. Guttikonda, H. MohamadAbbas, K. H. K, S. S. Kumar and R. K. Gupta, "Student Performance Prediction for Adaptive Learning using Improved COOT Optimization Algorithm," 2025 International Conference on Intelligent Computing and Knowledge Extraction (ICICKE), Bengaluru, India, 2025, pp. 1-6, doi: 10.1109/ICICKE65317.2025.11136713.
10. Nuvvula, Ramakrishna SS, Polamarasetty P. Kumar, Syed Riyaz Ahammed, R. Siva Subramanyam Reddy, Bachina Harish Babu, and Ahmed Ali. "Distributed Reinforcement Learning Framework for Autonomous Optimization of Grid-Scale Energy Storage Systems in Renewable Energy Integration." In 2024 12th International Conference on Smart Grid (icSmartGrid), pp. 290-297. IEEE, 2024.
11. S. B. Venkata, K. Thiyagarajan, S. C K, S. Thallal and M. Soni, "Improved Prototypical Network with Diacritic Detection for Few-Shot Urdu Handwritten Character Recognition," 2025 2nd International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS), Hassan, India, 2025, pp. 1-5, doi: 10.1109/IACIS65746.2025.11210962.
12. K. Thiyagarajan, Z. Sataar, Srinivasan, V. E and P. P. Selvam, "An Adaptive Multilingual and Morphology-Aware Framework for Low-Resource Indian Language Translation," 2025 International Conference on Intelligent Communication Networks and Computational Techniques (ICICNCT), Bidar, India, 2025, pp. 1-6, doi: 10.1109/ICICNCT66124.2025.11232851.
13. R. V. Durga, S. Manne, R. Vaddi, D. Paturu and D. S. S. Pavani, "IoT-Based Soil Moisture Monitoring with Remote Sensing validation for Precision agriculture," 2026 IEEE International Conference on AI Engineering and Innovations (AIEI), NIT Jamshedpur, India, 2026, pp. 1-6, doi: 10.1109/AIEI69164.2026.11496993.
14. Bhanu sekhar Guttikonda, Rakesh Guduru, Reddaiah Kasturi, Raghavendar Nellikondi, Shanmukha Rao Nadella, "Integrated Image Dataset for CI/CD, MLOps, and Low-Code AI Experiments", IEEE Dataport, November 1, 2025, doi:10.21227/wm2t-h572
15. Kumar, Polamarasetty P., Akhlaqur Rahman, Ramakrishna SS Nuvvula, Ilhami Colak, S. M. Muyeen, Sk A. Shezan, G. M. Shafiullah et al. "Using energy conservation-based demand-side management to optimize an off-grid integrated renewable energy system using different battery technologies." Sustainability 15, no. 13 (2023): 10137.
16. Ponniran, Asmarashid Bin, Ashok Vajravelu, Wan Suhaimizan Bin Wan Zaki, T. Yamunarani, Syed Riyaz Ahammed, and S. Sivaranjani. "IoT/WSN-based Security/privacy Methods for Forest Monitoring." In 2024 International Conference on Future Technologies for Smart Society (ICFTSS), pp. 146-151. IEEE, 2024.
17. AS. Vickram and B. B. Sowndharya, 'Breast histology slides classification using fused lightweight deep features', in 2024 5th International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2024, pp. 1–5.
18. Rao, Ch Hari Govinda, Ahmed Ali, Kush Patel, Syed Riyaz Ahammed, and Dillip Kumar Sahoo. "Performance Analysis of Smart Grid Energy Systems Under Varying Loads Using Digital Twin and Real-Time Data Analytics." In 2025 13th International Conference on Smart Grid (icSmartGrid), pp. 383-389. IEEE, 2025.