Bol. Soc. Paran. Mat.
©SPM – E-ISSN-2175-1188
SPM: www.spm.uem.br/bspm

(3s.) **v. 2026 (44)** : 1–15.
ISSN-0037-8712
doi:10.5269/bspm.81550

Integrating Bootstrap DEA and Machine Learning Meta-Frontiers to Assess Technical Efficiency and Technology Gaps in Gherkin, Paddy and Groundnut Farming Systems in Northern Tamil Nadu

A. Srikrishna and G. Y. Mythili

ABSTRACT: This paper analyses technical efficiency and technology gaps among 305 smallholder farmers cultivating gherkin, paddy, and groundnut in northern Tamil Nadu using a dual-frontier framework. Crop-specific technical efficiency is estimated via input-oriented Data Envelopment Analysis under variable returns to scale with bootstrap bias correction. To account for technology heterogeneity across crops, a DEA meta-frontier is constructed and Technology Gap Ratios are derived. In parallel, Quantile Random Forests are employed to estimate flexible nonparametric production frontiers at the 95th conditional quantile, and a pooled machine-learning meta-frontier is obtained. Results indicate high technical efficiency across all crops, with paddy and groundnut operating closer to their group frontiers, while gherkin exhibits greater dispersion. Meta-frontier estimates show minimal technology gaps, with gherkin technology lying closest to the global frontier. Strong agreement between DEA- and ML-based frontiers confirms the robustness of the findings.

Keywords: Technical efficiency, DEA bootstrap, meta- frontier, random forest, machine learning, gherkin, paddy, groundnut, Tamil Nadu.

Contents

2020 *Mathematics Subject Classification:* 90C05, 62G08, 68T05, 62P20.
Submitted January 25, 2026. Published April 17, 2026.

1. Introduction

Agriculture remains a cornerstone of the rural economy in Tamil Nadu, where smallholder farmers continue to supply most of the state's food and a growing share of high-value crops. Despite improvements in irrigation, market access and infrastructure, smallholders face tightening constraints from rising input prices, labour scarcity, climatic shocks and ongoing land fragmentation [1]. These pressures make it increasingly important not only to raise productivity, but to improve input-use efficiency and align cropping choices with technologies that can sustain farm incomes under volatile conditions [1, 4].

Northern Tamil Nadu illustrates these challenges in a particularly interesting way. Within the same agro-climatic setting, farmers allocate land among paddy, groundnut, and gherkin. Paddy remains the staple food crop, supported by canal and groundwater irrigation and long-established production practices. Groundnut is a key oilseed grown mostly under semi-arid conditions and exposed to rainfall risk. Gherkin, in contrast, is produced as a high-value export vegetable, primarily under contract farming arrangements in which companies provide seeds, technical advice and assured procurement [2, 3]. Recent evidence from South Asia shows that such contracts can reshape input use, risk management and technology adoption among smallholders [8, 9, 14].

Empirical work on rice in India and other Asian countries consistently finds sizeable technical inefficiencies, often associated with sub-optimal water management, delayed field operations and imperfect fertiliser use [10, 11, 12]. In South Indian paddy systems, studies emphasise how irrigation reliability, mechanisation and varietal choice jointly influence efficiency [5, 11]. Groundnut and other oilseeds also exhibit persistent efficiency gaps, with farmers constrained by poor access to improved seed, pest management services and timely extension support [13]. Oilseed performance is further exposed to climate variability, which compounds yield risk and accentuates the role of technology and adaptation strategies [1, 6].

In contrast, recent work on contract-farmed and high-value crops suggests that stronger integration with markets and input suppliers can lead to higher technical efficiency [8, 9, 14]. Studies on vegetables, fruits and livestock under formal contracts report that participating farms often operate closer to the production frontier than non-participants, although the magnitude of this advantage depends on contract design, monitoring and support services [7, 15, 16].

Evidence from Vietnam and Ethiopia, for example, indicates that contract farming can tighten the distribution of technical efficiency and reduce dispersion in performance across households [17, 18]. Yet, higher profitability under contract farming does not automatically reveal whether performance gains stem from better technology, more efficient input allocation or favourable pricing and risk-sharing arrangements. The case of gherkin in northern Tamil Nadu is a striking example. Farmers typically cultivate gherkin on small land parcels, but often report higher gross returns and profits than those growing paddy or groundnut.

Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA) are the two workhorses in this field, and remain widely used in recent Scopus- and Web of Science-indexed studies on smallholder production. DEA is particularly attractive in multi-input, single-output settings like small farms, where it is difficult to specify a parametric production function. However, classical DEA is deterministic and sensitive to sampling variation, which can bias efficiency scores. Bootstrap DEA procedures that allow bias correction and statistical inference, and these have been applied in numerous recent agricultural studies, including sugarcane, dairy, maize and beef cattle systems [20, 27, 28]. A further complication arises because gherkin, paddy and groundnut are produced under heterogeneous technologies. Estimating one common frontier for all crops would ignore these differences; estimating separate frontiers would prevent cross-crop comparisons. The meta-frontier framework, addresses this by modelling a global frontier that envelops crop-specific group frontiers and allows a Technology Gap Ratio (TGR) to be computed for each group. Meta-frontiers have been used to compare technologies across dairy farms in Europe, rice systems in Asia, horticultural farms in Africa and, more recently, contract versus non-contract livestock producers [18, 19, 20, 29].

At the same time, ML has emerged as a powerful tool for modelling complex, nonlinear production processes. Recent reviews show a rapid expansion of ML applications in crop yield prediction, input optimisation and risk assessment [23, 24]. Random Forest (RF) and its extensions are particularly popular due to their robustness and high predictive accuracy. QRF extend RF to model conditional quantiles, allowing estimation of high-output frontiers rather than just mean relationships. QRF has been used for probabilistic crop yield forecasting and for identifying yield potential and yield gaps under climate variability [24, 25, 26]. Despite these advances, there are still very few applications that jointly use bootstrap DEA with meta-frontiers and ML-based frontier estimation in smallholder agriculture, especially in multi-crop settings within a single region. This study addresses that gap by analysing technical efficiency and technology gaps among 305 smallholder farmers cultivating gherkin, paddy and groundnut in northern Tamil Nadu. We implement a dual-frontier framework, such as input-oriented, bias-corrected DEA models are estimated separately for each crop, and a pooled DEA meta-frontier is constructed to derive crop-level TGRs are from Econometric-operations research side, crop-specific 95th-quantile Random Forest frontiers are estimated, and a pooled QRF meta-frontier is used to obtain ML-based efficiency and technology gap measures are from Machine-learning side. Combining these approaches yields a richer view of performance such as DEA provides transparent, theory-consistent measures of technical efficiency, while QRF captures nonlinearities and complex interactions in the production process. This allows us to examine whether gherkin's apparent advantage reflects higher efficiency, superior technology, or a mixture of both, and how far paddy and groundnut farmers could move towards the global best-practice frontier.

The remainder of the paper is structured as follows. Section 2 reviews the relevant literature on efficiency analysis, meta-frontier modelling, and machine-learning frontiers. Section 3 describes the methodological framework, including the bootstrap DEA approach and the ML based quantile Random Forest frontier. Section 4 presents the empirical results and discusses the efficiency patterns and technology gaps across crops. Section 5 concludes with key findings and policy implications.

2. Literature Review

2.1. Technical Efficiency in Smallholder Crop Systems

Technical efficiency analysis has been a core theme in agricultural economics since Farrell's seminal work and continues to guide current empirical research. Recent studies across Asia and Africa consistently show that smallholder farmers operate below the production frontier, especially in cereal- and legume-based systems. For rice in South and Southeast Asia, empirical evidence links inefficiencies to irrigation unreliability, labour shortages, late transplanting and sub-optimal fertiliser and pesticide use. Improvements in water management and mechanisation are repeatedly identified as effective pathways to reduce these gaps, as highlighted in studies on India, China and the Indo-Gangetic Plains [10, 11, 12].

For oilseeds such as groundnut, parallel evidence emerges. Research in India, Ghana and Nigeria shows that inefficiency is strongly associated with seed quality constraints, pest management challenges, limited credit access and weak extension support. Climate variability and rainfall shocks are especially important in semi-arid environments, further depressing groundnut yields and emphasising the need for improved varieties and climate-resilient agronomic strategies [1, 6, 13].

High-value horticultural crops have also attracted increasing research attention as diversification accelerates in emerging economies. Studies from Ethiopia, India and other developing countries show that horticultural farmers often achieve comparatively high technical efficiency, particularly when they are integrated into organised supply chains or contract farming arrangements. For example, a 2025 study of vegetable producers in Ethiopia reports moderate-to-high technical efficiency but identifies substantial potential for improvement through better input allocation and enhanced training and advisory services [17, 18].

2.2. DEA, DEA Bootstrap and Meta-Frontiers

DEA remains one of the most widely used tools to evaluate farm-level efficiency, and new applications continue to appear in Scopus- and Web of Science-indexed journals across crops and regions. Recent work on pineapple farmers in Ghana, sugarcane producers in Indonesia, and maize and dairy farms elsewhere shows that DEA is particularly useful for multi-input systems where parametric forms are difficult to specify [32, 33, 36]. However, the traditional DEA estimator is deterministic and can overstate efficiency due to sampling noise. Building on earlier theoretical contributions, Simar and Wilson proposed bootstrap DEA procedures that adjust for bias and provide confidence intervals [30, 31], and these have become standard in modern DEA applications. Recent agricultural studies-ranging from sugarcane in East Java to beef cattle and dairy farms-apply bootstrap DEA or double-bootstrap DEA to obtain more reliable efficiency measures and to study determinants of inefficiency [27, 28, 33]. The meta-frontier framework further enriches efficiency analysis by accommodating technology heterogeneity. Under this framework, groups (e.g., different crops or regions) each have their own frontier, and a global meta-frontier envelops these. The Technology Gap Ratio (TGR) quantifies how far a group's technology lies below the meta-frontier. Meta-frontier models have been applied in dairy production, crop systems and environmental efficiency analysis [34, 35, 29], and more recently to compare contract and non-contract farms [18, 19, 20]. A recent example is Huong et al. (2025), who use a DEA meta-frontier to compare contract and non-contract pig farms in Vietnam, showing that contract farms are more efficient under their group frontier but may still exhibit technology gaps relative to the meta-frontier [18].

2.3. Machine Learning and Quantile Frontiers in Agriculture

Machine learning has rapidly gained ground in agricultural research, with applications spanning yield prediction, disease detection, irrigation optimisation and farm management support [23, 24, 38]. Random Forests in particular have been widely used because they handle nonlinear relationships and interactions without demanding extensive prior assumptions. Recent studies from Ghana, India and global datasets show that RF performs strongly in predicting yields and agronomic efficiency [25, 26]. Quantile Random Forest (QRF) extends RF to estimate conditional quantiles of the output distribution. In the context of production, upper quantiles can be interpreted as data-driven frontiers representing attainable yields or revenues given input levels. Recent work uses QRF and related methods to produce probabilistic forecasts of maize and wheat yields, assess the effects of climate shocks, and identify yield gaps [23, 24, 25]. Studies also show that QRF tends to outperform quantile neural networks in quantile regression tasks related to crop production [26]. Although still emerging, ML-based frontier and ML meta-frontier approaches have started to appear in energy, environmental and agricultural efficiency contexts. For example, recent work uses ML-based meta-frontiers to examine environmental efficiency differences across regions or technologies, showing that flexible algorithms can capture complex heterogeneity that traditional models may miss [29, 38].

2.4. Hybrid DEA-ML Approaches

Finally, there is a growing recognition that combining DEA with ML can yield richer insights than using either method alone [39, 40, 41, 42]. Recent reviews highlight hybrid frameworks where DEA scores are used as targets in ML models, ML is used to approximate frontiers, or both approaches are employed side-by-side to cross-check robustness [39, 40]. Such hybrid approaches are increasingly applied in energy and industrial production, but still relatively rare in smallholder agriculture [41, 42]. Given this gap, a combined bootstrap DEA-meta-frontier and QRF meta-frontier approach, as proposed in this study, is both novel and well aligned with emerging methodological trends.

3. Methodology

This study develops a dual-frontier framework to assess technical efficiency and technological heterogeneity among smallholder gherkin, paddy and groundnut farmers in northern Tamil Nadu. The framework combines an econometric–operations research frontier with a machine-learning-based frontier as non-parametric production frontier.

3.1. Data and Production Variables

The analysis is based on cross-sectional data from 305 farmers, each cultivating one of three crops-gherkin, paddy, or groundnut-during a single agricultural year in Northern Tamil Nadu, particularly in the districts of Tiruvannamalai, Villupuram, and Kallakurichi. For every farmer, a set of input and output variables was compiled. Inputs include the land area devoted to the crop (measured in cents), expenditure on fertilisers and pesticides, spending on machinery and hired labour, and the total input cost covering all purchased inputs. The output variable is the total revenue earned from the crop, expressed in rupees. Using monetary measures for inputs and output enables the estimation of efficiency in economic terms and aligns with recent DEA studies in crop and livestock production [6, 7, 27, 32]. This choice is also appropriate in a policy context where the primary concern is farm income and profitability rather than only physical yields [4, 5, 21]. Because gherkin, paddy and groundnut differ considerably in their agronomic requirements, production environments and market structures, each crop is modelled as a separate technology group in the subsequent analysis [2, 3, 13]. All analyses were carried out in R software, which was used to implement both the DEA and ML-based frontier models [23, 25, 26, 39].

3.2. Data Envelopment Analysis (DEA)

Technical efficiency within each crop group is first estimated using input-oriented DEA under variable returns to scale (VRS). DEA constructs a non-parametric, piecewise-linear frontier by enveloping observed input-output combinations. A farm is considered efficient if it lies on the frontier; otherwise, efficiency is measured by the proportional reduction in inputs required to reach the frontier while keeping output constant.

For each crop, let each farmer be a decision-making unit (DMU) using input vector x_i to produce output y_i. The VRS input-oriented DEA model solves, for each farm i:

$$\min_{\theta_i,\lambda_j} \quad \theta_i \tag{3.1}$$

$$\text{subject to} \quad y_i \leq \sum_j \lambda_j y_j \tag{3.2}$$

$$\theta_i x_i \geq \sum_j \lambda_j x_j \tag{3.3}$$

$$\sum_j \lambda_j = 1, \quad \lambda_j \geq 0 \tag{3.4}$$

where θ_i is the technical efficiency score and λ_j are weights forming convex combinations of peer farms. A value $\theta_i = 1$indicates full efficiency; values below one indicates the proportion by which inputs could be reduced. This specification follows standard agricultural DEA practice [6, 7, 10, 12, 13]. Separate DEA models are estimated for gherkin, paddy and groundnut to obtain group-specific efficiency scores and frontiers.

3.3. Bootstrap DEA

Classical DEA estimates are sensitive to sampling noise and tend to be biased upwards because the estimated frontier is treated as if it were known. To address this, we apply the bootstrap DEA procedure proposed by Simar and Wilson [30, 31]. For each crop group, Compute the original DEA efficiency scores $\hat{\theta}_i$. Generate B bootstrap pseudo-samples via resampling residuals and re-estimating the frontier (here B=200). For each pseudo-sample b, compute efficiency scores $\hat{\theta}_i^{(b)}$.

Estimate the bias as,

$$\text{Bias}(\hat{\theta}_i) = \frac{1}{B} \sum_{b=1}^{B} \left(\hat{\theta}_i^{(b)} - \hat{\theta}_i \right) \tag{3.5}$$

Obtain the The bias-corrected score,

$$\hat{\theta}_i^{BC} = \hat{\theta}_i - \text{Bias}(\hat{\theta}_i) \tag{3.6}$$

3.4. DEA Meta-Frontier and Technology Gap Ratio

Because the three crops operate under different technologies, their DEA efficiency scores are not directly comparable. To evaluate how each crop's technology performs relative to the best available technology in the sample, we estimate a DEA meta-frontier, following the meta-frontier literature of Battese and later DEA applications [19, 20, 29, 34]. A pooled VRS input-oriented DEA is estimated for all 305 farmers, producing meta-frontier efficiency $\hat{\phi}_i$ for each farm i. The Technology Gap Ratio (TGR) is then defined as:

$$\text{TGR}_i = \frac{\hat{\theta}_i^{BC}}{\hat{\phi}_i} \tag{3.7}$$

where $\hat{\theta}_i^{BC}$ is the bias-corrected group efficiency, and $\hat{\phi}_i$ represents the meta-frontier efficiency. A TGR close to one implies that the group's technology is close to the meta-frontier; lower TGR values indicate a more pronounced technology gap. Crop-level average TGRs provide a measure of how gherkin, paddy and groundnut technologies compare to the global best-practice frontier in this region.
This approach follows the meta-frontier literature pioneered by Battese and co-authors and extended in many recent DEA applications, including contract farming and environmental efficiency studies.

3.5. Quantile Random Forest Frontier (ML Frontier)

To complement DEA and allow for nonlinearities, we estimate machine-learning frontiers using Quantile Random Forests. For each crop, we fit a QRF model with total revenue as the response and the four inputs (land, fertiliser-pesticide cost, machinery-labour cost, total input cost) as predictors. Random Forests consist of an ensemble of regression trees trained on bootstrapped samples with random feature selection; QRF further aggregates tree predictions to provide conditional quantiles rather than mean predictions. For each farmer i, we obtain the predicted 95th conditional quantile of revenue $\hat{y}_i^{QRF}$, interpreted as the data-driven frontier revenue given that farmer's input combination.

The machine-learning–based technical efficiency is defined as:

$$TE_i^{ML} = \min\left(\frac{y_i}{\hat{y}_i^{QRF}}, 1\right) \tag{3.8}$$

where y_i is observed revenue. Values close to one indicate that a farmer is operating near the ML-estimated frontier. Recent studies demonstrate the suitability of RF and QRF for yield modelling, agronomic efficiency assessment and probabilistic forecasting under climate variability [23, 24, 25, 26, 38].

3.6. QRF Meta-Frontier and ML-Based Technology Gaps

To mirror the DEA meta-frontier in the ML framework, we estimate a pooled QRF model using all farmers from the three crops. Again, the 95th conditional quantile of revenue is predicted for each observation, yielding ML meta-frontier values $\hat{y}_i^{META}$.

The corresponding ML meta-efficiency is:

$$TE_i^{(ML,META)} = \min\left(\frac{y_i}{\hat{y}_i^{META}}, 1\right) \tag{3.9}$$

The machine-learning Technology Gap Ratio is given by:

$$\mathrm{TGR}_i^{ML} = \frac{TE_i^{ML}}{TE_i^{(ML,META)}} \tag{3.10}$$

These measures capture how far each crop's effective technology lies from the ML-estimated global frontier, providing a nonlinear, data-driven counterpart to the DEA-based TGRs. This approach is inspired by recent work applying ML algorithms to meta-frontier-type questions in environmental and operational efficiency [39, 42, 43, 44].

3.7. Integrating DEA and ML Results

Using both DEA and QRF frontiers allows us to triangulate efficiency and technology gap patterns. If both methods indicate high efficiency and small technology gaps for gherkin relative to paddy and groundnut, this strengthens the argument that contract-farmed gherkin enjoys a genuine technological and managerial advantage [2, 8, 14, 22]. If DEA suggests small gaps but QRF reveals nonlinearities that favour certain input bundles (e.g., small land–high input intensity combinations), this may highlight the role of crop-specific biological responses or contract design. If results diverge systematically, that itself is informative about the limitations of linear frontiers or the presence of strong nonlinear interactions [1, 4, 5].

By aligning a classical, interpretable frontier approach (bootstrap DEA + meta-frontier) with a flexible, modern ML approach (QRF frontiers + ML meta-frontier), this methodology is designed to provide robust insights into smallholder efficiency and technological heterogeneity in northern Tamil Nadu, and to inform policy choices around crop diversification, contract farming and technology dissemination.

4. Results and Discussion

This section presents the empirical findings from the dual-frontier framework and discusses their implications. We first examine crop-wise technical efficiency from bootstrap-corrected DEA, then evaluate technology gaps using the DEA meta-frontier. We next present the ML frontier results based on QRF and their ML meta-frontier. Finally, we compare the two approaches and interpret what the results imply for gherkin, paddy and groundnut farmers in northern Tamil Nadu.

Table 1: Cost and Returns of Sample Farmers by Crop (Mean and Standard Deviation)

Cost / Returns	Gherkin Mean (SD)	Paddy Mean (SD)	Groundnut Mean (SD)
Size of land (in cents)	30.9180 (12.932)	155.0918 (102.080)	153.5246 (96.497)
Input cost	3032.7869 (1238.994)	2323.4426 (1531.309)	9209.7377 (5787.159)
Fertilizer and pesticide	18554.0984 (7736.482)	9576.0984 (6195.331)	9761.2984 (6135.981)
Machinery and labour	6406.2295 (2664.779)	15150.4918 (9731.285)	11815.7574 (9215.664)
Total revenue	82599.8361 (34483.067)	111475.9016 (73288.065)	118023.3443 (74804.504)

Table 1 represents mean standard deviation values of cost and returns of farmers for all three crops. Table 1 indicates that gherkin generates substantially higher revenue per unit of land. Gherkin fertilizer cost is heavy than Paddy and groundnut. likely machinery and labour cost less than those two. Groundnut input cost higher than comparing to other two crops.

4.1. Technical efficiency from bootstrap DEA

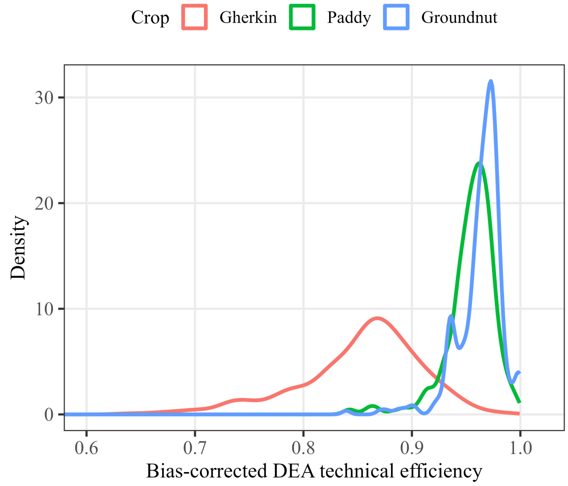
Table 2 reports the group-specific DEA efficiency scores for the three crops. The first column shows the mean VRS, input-oriented DEA efficiency, while the second column presents the bias-corrected efficiency obtained via the Simar–Wilson bootstrap procedure [30, 31].

Table 2: Mean group-specific technical efficiency scores (DEA and bootstrap-corrected)

Crop	Mean DEA efficiency	Mean DEA Bootstrap efficiency	DEA Bootstrap SD	DEA Bootstrap Min	DEA Bootstrap Max
Gherkin	0.865	0.850	0.067	0.384	0.986
Paddy	0.959	0.952	0.041	0.353	0.997
Groundnut	0.965	0.960	0.039	0.562	0.999

As shown in Table 2, the bootstrap correction reduces the estimated efficiency levels slightly for all crops, confirming that conventional DEA tends to overestimate efficiency because it treats the estimated frontier as if it were known and deterministic. The downward adjustment is modest (around 0.01-0.02 points), indicating that while sampling noise is present, the underlying data are reasonably consistent and not dominated by extreme outliers, in line with findings from recent agricultural applications of bootstrap DEA.

Figure 1. Distribution of bias-corrected DEA technical efficiency by crop



Two clear patterns emerge. First, paddy and groundnut farmers operate close to their respective group-specific frontiers. The mean bias-corrected efficiency is 0.952 for paddy and 0.960 for groundnut, suggesting that, on average, farmers in these crops could reduce inputs by only about 4-5% without sacrificing output. Similar high efficiency levels have been reported for staple cereals and oilseeds when farmers have reliable access to irrigation, improved varieties and basic mechanisation, for instance in Indian rice and groundnut systems [11, 12, 13]. These results are therefore consistent with the idea that paddy and groundnut technologies in the study area are mature and well internalised by farmers. Second Gherkin farmers show lower average technical efficiency (0.850). In contrast, gherkin farmers could theoretically reduce inputs by about 15% at current output levels if they operated on their crop frontier. This does not imply that gherkin is a “worse” crop. Rather, it reflects greater heterogeneity in management quality and input-use across farmers, which is plausible for an intensive horticultural crop requiring more frequent decisions on pest management, irrigation timing and labour allocation [2, 3, 22]. Studies on export-oriented vegetables and high-value horticulture often find higher variability in technical efficiency, even when average profitability is high. These patterns are clearly visualised in Figure 1, where gherkin efficiency scores exhibit a wider and more dispersed distribution compared with the sharply peaked distributions of paddy and groundnut. The relatively tight clustering of efficiency scores for paddy and groundnut further supports the interpretation that these technologies are well established and consistently applied among farmers [6, 7]. In contrast, the broader spread of gherkin scores indicates managerial and operational differences that leave considerable room for efficiency improvement within the crop group.

4.2. DEA meta-frontier and Technology Gap Ratio (TGR)

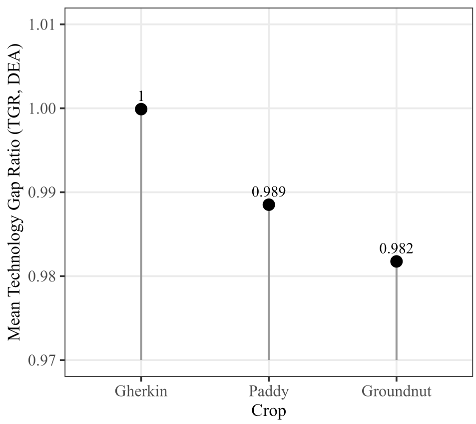
Because gherkin, paddy and groundnut differ in agronomy and institutional arrangements, farmers are assumed to operate under distinct production technologies. Group-specific DEA scores cannot therefore be compared directly. The meta-frontier framework resolves this by estimating a common enveloping frontier across all groups, allowing technology differences to be summarised using the Technology Gap Ratio (TGR).
A pooled VRS, input-oriented DEA model was estimated using the combined sample of 305 farmers. From this, meta-frontier efficiency scores and TGR were computed for each crop (Table 3 & Figure 2).

The TGR compares each crop’s group frontier with the meta-frontier. A value of 1 indicates that the group technology lies exactly on the meta-frontier, while values less than 1 indicate that the group technology is inferior to the global best practice [2, 9, 18].

Table 3: DEA meta-frontier efficiency and Technology Gap Ratio (TGR)

Crop	Mean meta-frontier efficiency	Mean group efficiency (bias-corrected)	Mean TGR	TGR SD	TGR Min	TGR Max
Gherkin	0.865	0.850	0.999	0.001	0.983	1.000
Paddy	0.945	0.952	0.989	0.011	0.951	1.000
Groundnut	0.944	0.960	0.982	0.008	0.944	1.000

Figure 2. DEA-based Technology Gap Ratio (TGR) by crop



Three key insights arise, first one is Gherkin technology coincides with the meta-frontier (TGR= 0.999). The result that gherkin's TGR 0.999 implies that gherkin farmers, as a group, have access to the most advanced production technology in the sample. This aligns with the institutional reality of gherkin being produced under contract farming: companies typically supply hybrid seeds, recommended input packages and technical advice, and impose quality standards, which effectively push the production technology toward a high, standardised frontier. Gherkin's lower efficiency (Section 4.1) is therefore not due to poor technology, but due to how that technology is used at the farm level. Second, paddy and groundnut technologies lie marginally below the global meta-frontier [4, 12, 29]. TGR values of 0.989 for paddy and 0.982 for groundnut suggest small but non-negligible technology gaps relative to gherkin. This is consistent with the broader literature where staple and oilseed crops, though important, tend to receive more incremental innovations and less intensive private-sector support compared to export-oriented horticulture [20, 34]. The implication is that paddy and groundnut farmers may be operating with slightly older or less intensive technologies-such as less advanced varieties, weaker integration with advisory services, or suboptimal irrigation systems. Last one is technology gaps are numerically small but economically relevant. While all TGR values are close to 1, even small differences can matter over time, especially in a context of rising input prices and climate risk. Previous meta-frontier studies have similarly found that modest technology gaps across groups can have significant long-run productivity and income implications.

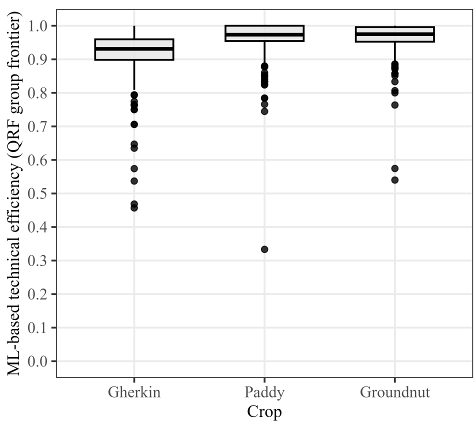
4.3. Machine-learning frontiers (QRF) at group level

To capture potential nonlinearities and complex input interactions beyond the linear piecewise DEA frontier, the study estimated crop-specific frontiers using QRF. QRF is particularly suitable because it extends Random Forests to directly model conditional quantiles, allowing the 95th percentile of revenue to be interpreted as a flexible, non-parametric frontier [23, 26, 38]. Table 4 summarises the mean QRF-based technical efficiency (TE) scores for each crop, computed as the ratio of observed revenue to the predicted 95th-quantile revenue, truncated at one.

Table 4: QRF-based group frontier efficiency scores

Crop	Mean TE (ML group frontier)	ML group frontier SD	ML group frontier Min	ML group frontier Max
Gherkin	0.920	0.076	0.457	1.000
Paddy	0.964	0.055	0.333	1.000
Groundnut	0.965	0.049	0.540	1.000

Figure 3. ML-based technical efficiency (QRF group frontier) by crop



This strong alignment between DEA and ML results supports the robustness of the findings and echoes recent studies where Random Forest-based frontiers performed very similarly to non-parametric or semi-parametric econometric frontiers in agriculture and energy [37, 39, 41].

Figure 3 shows that slightly higher efficiency of gherkin under QRF than under DEA (0.92 vs 0.85) suggests that the ML frontier is better able to accommodate nonlinear patterns and interactions involving land, input cost and other variables. For instance, the fact that gherkin is often cultivated intensively on small plots may generate a complex response surface where marginal gains from additional inputs are not linear [38]. ML methods like QRF can flexibly "bend" the frontier to fit regions of dense data, which can result in somewhat higher efficiency scores for farmers operating in those regions.

4.4. ML-based meta-frontier and ML Technology Gap Ratio (ML–TGR)

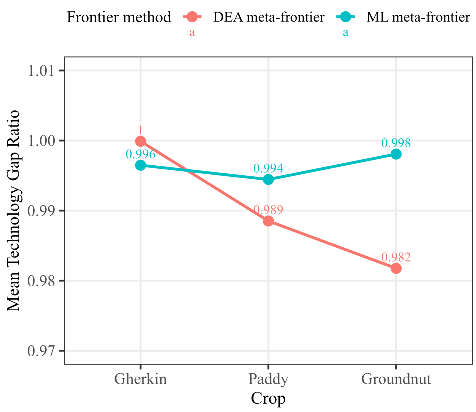
The DEA meta-frontier analysis was mirrored in the ML framework by estimating a pooled QRF model on the full sample, thus yielding an ML meta-frontier. Using this, ML meta-frontier technical efficiency and the ML Technology Gap Ratio were computed as ratios of group and meta efficiency [40].

Table 5: ML meta-frontier efficiency and ML Technology Gap Ratio

Crop	Mean TE (ML meta frontier)	Mean TE (ML group frontier)	Mean ML–TGR
Gherkin	0.917	0.920	0.996
Paddy	0.960	0.964	0.994
Groundnut	0.967	0.965	0.998

As shown in Table 5 & Figure 4, Three observations stand out. First observation is ML-TGR values are very close to 1 for all crops. This confirms that, even from a fully nonlinear ML perspective, technology gaps between the three crop groups are minimal [13, 21]. This result resonates with recent ML-based meta-frontier applications that find relatively small technology gaps once flexible models are used, particularly when groups share similar agro-ecological conditions and access to broad input markets. Next observation Groundnut shows the highest ML-TGR (0.998). This suggests that, according to the QRF meta-frontier, groundnut farmers' technology is extremely close to the global best-practice technology. This may reflect the fact that groundnut is grown in semi-arid zones where more intensive adoption of improved varieties and associated agronomy has recently taken place. It may also reflect the ML model's interpretation of the input-output relationships in the groundnut data. And last Observation Gherkin retains a slight technology edge in DEA but not in ML. In DEA, gherkin sits exactly on the meta-frontier (TGR = 1.0), whereas under QRF, its ML-TGR is 0.996, marginally below groundnut. Given the very small difference, this discrepancy should not be overinterpreted [39, 40, 44]. It mainly illustrates that different frontier estimators can produce slightly different envelopes when data points are tightly clustered near the frontier.

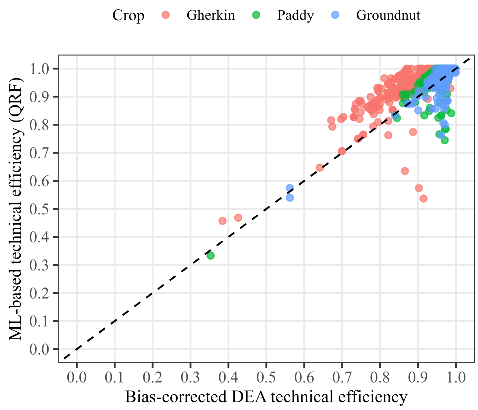
Figure 4. Comparison of DEA and ML meta-frontier Technology Gap Ratios



4.5. Comparing DEA and ML frontiers

From Figure 5, The combination of DEA (with bootstrap and meta-frontier) and QRF (with ML meta-frontier) allows us to cross-check the stability of efficiency patterns. Several insights follow from this comparison like Robust ranking across methods; Efficiency differences dominate technology differences; Nonlinearity and heterogeneity are present but not overwhelming. From Robust ranking across methods, both approaches identify gherkin as the relatively less efficient crop in technical terms, and paddy and groundnut as comparatively more efficient. This robustness across very different modelling strategies-linear programming versus ensemble tree-based learning-adds credibility to the empirical findings and follows a growing strand of literature that advocates hybrid DEA-ML analysis for productivity and efficiency studies [39, 42]. From Efficiency differences dominate technology differences, DEA TGR and ML-TGR values are uniformly close to 1, suggesting that the three crops share broadly similar technology levels given the input-output combinations and the local environment. Performance differences are therefore mainly efficiency-driven (how well farmers use available technology) rather than technology-driven (what technology they have). Similar conclusions have been drawn in meta-frontier studies comparing contract and non-contract farms, irrigated and rainfed systems, or regions with similar agro-ecology but different management cultures [8, 17, 19]. From Nonlinearity and heterogeneity are present but not overwhelming, ML allows for complex interactions and nonlinearities. The fact that QRF and DEA results are broadly consistent suggests that, in this dataset, the production function is not highly irregular; DEA's piecewise linear envelope is already a reasonable approximation. ML mainly refines the shape of the frontier and gives more nuanced information on dispersion, rather than overturning the efficiency ranking. This is consistent with findings in yield prediction studies where Random Forests improve predictive accuracy but often corroborate the main patterns found via simpler models [23, 26].

Figure 5. Relationship between DEA and ML-based technical efficiency



A notable finding from the analysis is that gherkin farmers, despite operating on much smaller plots, earn considerably higher revenue and profit per unit of land than those cultivating paddy or groundnut. The frontier estimates help explain this pattern. Although gherkin growers exhibit slightly lower technical efficiency on average (0.85-0.92), their Technology Gap Ratios are effectively equal to one, indicating that they operate with one of the most advanced technologies available among the three crop systems [2, 8, 22]. More importantly, gherkin is intrinsically a high-value crop embedded in a structured contract farming system that provides hybrid seeds, bundled inputs, technical guidance and guaranteed procurement at predetermined prices. These institutional features create strong economic advantages independent of pure technical efficiency. In essence, gherkin farmers benefit from participating in a profitable value chain, even if they have not completely mastered the production technology. This result aligns with broader evidence from high-value horticulture and contract farming studies, which commonly report higher incomes alongside significant inefficiency gaps among participating farmers [4, 11]. In contrast, paddy and groundnut farmers operate with relatively high technical efficiency (around 0.95–0.96) and only modest technology gaps, suggesting that they utilise their available technologies effectively. However, the technologies themselves are less profitable and more vulnerable to water scarcity, climatic variability and input price fluctuations [21]. This combination-high efficiency, small technology gaps, but modest returns-is consistent with a risk-averse production strategy in which smallholders maintain familiar staple or semi-arid crops for security while cautiously integrating high-value crops like gherkin to enhance household income. From a policy perspective, these findings justify differentiated interventions: reinforcing management and extension support in gherkin to close efficiency gaps; and focusing on technological upgrading and diversification options for paddy and groundnut farmers who are already using their existing technologies relatively efficiently.

5. Conclusion

This study assessed the technical efficiency and technological diversity of 305 smallholder farmers growing gherkin, paddy, and groundnut in Tamil Nadu. A dual frontier approach combining DEA Bootstrap with ML-based QRF frontiers was used. The results of this study provide a comprehensive empirical picture of efficiency differences across cropping systems. The bootstrap-corrected DEA results show that all three crops show relatively high levels of technical efficiency, with gherkin farmers at 85% efficiency and paddy and groundnut farmers at more than 95%. These results suggest that despite differences in cropping environments, resource endowments, and institutional arrangements, farmers are generally efficient in converting available inputs into output. It further reveals that technological diversity exists in all crops but not on a large scale. The TGR values of all crops are above 98%. The Gherkin shows very close alignment with the meta-frontier. This indicates that the overall level of technological progress across farmers is relatively similar, even though each crop operates under a unique production technology. However, horticultural systems demonstrate a small technological edge. ML-based QRF frontiers confirm the DEA findings, producing stable performance rankings and similarly narrow technological gaps. The

ML meta-frontier suggests that nonlinearities and complex production interactions slightly reduce the perceived technological gaps. This suggests that some of the constraints observed under DEA may reflect model structure rather than true technological differences. Gherkin farmers were more profitable despite having smaller landholdings. This reflects the transformative role of contract farming. which is characterized by integrated input supply, consistent extension support, and guaranteed purchases. It also improves efficiency, reduces market risk, and improves compliance with production protocols. In contrast, rice and groundnut farmers are technically efficient within their resource environments. The technologies they use seem to have little ability to push production towards the global frontier. The findings highlight the usefulness of combining traditional frontier analysis with machine learning methods to identify performance in multi-crop smallholder systems. The dual-method approach provides a more nuanced understanding of how production technologies, organizational arrangements, and crop characteristics collectively shape farmer performance. Future research could extend this framework by integrating environmental performance metrics, analyzing multi-output systems and combining group data to explore performance patterns that change over time.

References

1. V. Touch, D. K. Tan, B. R. Cook, D. L. Liu, R. Cross, T. A. Tran and A. Cowie, *Smallholder farmers' challenges and opportunities: Implications for agricultural production, environment and food security*, Journal of Environmental Management 370 (2024), 122536. https://doi.org/10.1016/j.jenvman.2024.122536.

2. K. C. Gummagolmath and P. Sharma, *Leveraging contract farming for supply chain in India: A case of gherkin production*, Journal of Agricultural Extension Management 16 (2018), no. 1, 11–25.

3. H. D. Venu Prasad, P. Singh and V. K. Chaturvedi, *Socio-economic impact analysis of contract farming of gherkin in Karnataka*, Indian Journal of Extension Education 48 (2012), no. 1–2, 42–44.

4. A. K. Basantaray, S. Acharya and T. Patra, *Crop diversification and income of agricultural households in India: An empirical analysis*, Discover Agriculture 2 (2024), no. 1, 8. https://doi.org/10.1007/s44279-024-00019-0.

5. R. J. Varadan, S. Mamidanna, S. Kumar, S. Z. Ahmed and I. Jaisankar, *Technology, infrastructure and enterprise trade-off: Strengthening smallholder farming systems in Tamil Nadu State of India for sustainable income and food security*, Outlook on Agriculture 51 (2022), no. 2, 197–212. https://doi.org/10.1177/00307270221077380.

6. G. A. Balaji, V. Geethalakshmi, A. Senthil, M. Prahadeeswaran, S. Iswarya, M. Rajavel and S. Priyanka, *Assessment of economic efficiency and its determinants for mixed crop–livestock production under dryland agriculture systems in Tamil Nadu, India*, Sustainability 15 (2023), no. 10, 8332. https://doi.org/10.3390/su15108332.

7. S. Rajendran, *Technical efficiency of fruit and vegetable producers in Tamil Nadu, India: A stochastic frontier approach*, Asian Journal of Agriculture and Development 11 (2014), no. 1, 77–93. https://doi.org/10.22004/ag.econ.200294.

8. K. R. Paltasingh and P. K. Jena, *Does contract farming enhance farm efficiency? A case of wheat growers of Haryana, India*, Heliyon 9 (2023), no. 4, e15222. https://doi.org/10.1016/j.heliyon.2023.e15222.

9. Q. Yang, X. Zhang, H. Peng and D. Sun, *Contract farming for improving farmers' incomes: From a resource-based view*, Frontiers in Sustainable Food Systems 9 (2025), 1585843. https://doi.org/10.3389/fsufs.2025.1585843.

10. I. Sharma and M. K. Sekhon, *Measurement and determinants of technical efficiency in crop production: A review*, Agricultural Reviews 43 (2022), no. 1, 77–83.

11. M. Mallareddy et al., *Maximizing water use efficiency in rice farming: A comprehensive review of innovative irrigation management technologies*, Water 15 (2023), no. 10, 1802. https://doi.org/10.3390/w15101802.

12. R. B. S. Chandel, A. Khan, X. Li and X. Xia, *Farm-level technical efficiency and its determinants of rice production in Indo-Gangetic plains*, Sustainability 14 (2022), no. 4, 2267. https://doi.org/10.3390/su14042267.

13. A. A. Reddy and M. C. S. Bantilan, *Competitiveness and technical efficiency: Determinants in the groundnut oil sector of India*, Food Policy 37 (2012), no. 3, 255–263. https://doi.org/10.1016/j.foodpol.2012.02.004.

14. A. Selorm et al., *Does contract farming affect technical efficiency? Evidence from soybean farmers in Northern Ghana*, Agricultural and Food Economics 11 (2023), no. 1, 9. https://doi.org/10.1186/s40100-022-00240-9.

15. A. G. Singbo, G. Emvalomatis and A. Oude Lansink, *The effect of crop specialization on farms' performance: A Bayesian non-neutral stochastic frontier approach*, Frontiers in Sustainable Food Systems 5 (2021), 711530. https://doi.org/10.3389/fsufs.2021.711530.

16. A. Gaviglio et al., *Technical efficiency and productivity of farms: A peri-urban case study analysis*, Agricultural and Food Economics 9 (2021), no. 1, 11. https://doi.org/10.1186/s40100-021-00181-9.

17. M. Yeshitila et al., *The impact of contract farming on technical efficiency in Ethiopia's smallholder sesame production*, Journal of the Austrian Society of Agricultural Economics 16 (2020), no. 4, 57–66.

18. L. T. T. Huong et al., *Does formal contract farming improve the technical efficiency of livestock farmers? A case study of fattening pig production in Hanoi, Vietnam*, Sustainability 17 (2025), no. 8, 3557. https://doi.org/10.3390/su17083557.

19. A. G. Geffersa and F. W. Agbola, *Technical efficiency and technology gaps among smallholder maize farmers in Ethiopia*, Sustainable Futures 9 (2025), 100620. https://doi.org/10.1016/j.sftr.2025.100620.

20. S. Aravindakshan et al., *Application of a bias-corrected meta-frontier approach and an endogenous switching regression to analyze the technical efficiency of conservation tillage for wheat in South Asia*, Journal of Productivity Analysis 49 (2018), no. 2, 153–171. https://doi.org/10.1007/s11123-018-0525-y.

21. M. J. Carrer et al., *Precision agriculture adoption and technical efficiency: An analysis of sugarcane farms in Brazil*, Technological Forecasting and Social Change 177 (2022), 121510. https://doi.org/10.1016/j.techfore.2022.121510.

22. A. K. Mishra et al., *Impact of contract farming on yield, costs and profitability in low-value crops*, Australian Journal of Agricultural and Resource Economics 62 (2018), no. 4, 589–607. https://doi.org/10.1111/1467-8489.12268.

23. T. van Klompenburg et al., *Crop yield prediction using machine learning: A systematic literature review*, Computers and Electronics in Agriculture 177 (2020), 105709. https://doi.org/10.1016/j.compag.2020.105709.

24. R. K. Sharma et al., *Maize and soybean yield prediction using machine learning methods: A systematic literature review*, Discover Agriculture 3 (2025), no. 1, 64. https://doi.org/10.1007/s44279-025-00215-6.

25. J. H. Jeong et al., *Random forests for global and regional crop yield predictions*, PLoS ONE 11 (2016), no. 6, e0156571. https://doi.org/10.1371/journal.pone.0156571.

26. M. Shahhosseini et al., *Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt*, Scientific Reports 11 (2021), no. 1, 1606. https://doi.org/10.1038/s41598-020-80820-1.

27. A. Karimov, *Productive efficiency of potato and melon growing farms in Uzbekistan: A two-stage double bootstrap data envelopment analysis*, Agriculture 3 (2013), no. 3, 503–515. https://doi.org/10.3390/agriculture3030503.

28. A. Charalampidou et al., *The power of cooperation: Evaluating dairy farm efficiency using bootstrap DEA*, Agriculture 15 (2025), no. 23, 2454. https://doi.org/10.3390/agriculture15232454.

29. H. Wang and L. Du, *Agricultural credit scale and agricultural green production efficiency: A metafrontier–Malmquist–Luenberger and panel Tobit approach*, Frontiers in Environmental Science 11 (2023), 1191012. https://doi.org/10.3389/fenvs.2023.1191012.

30. L. Simar and P. W. Wilson, *Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models*, Management Science 44 (1998), no. 1, 49–61. https://doi.org/10.1287/mnsc.44.1.49.

31. L. Simar and P. W. Wilson, *Estimation and inference in two-stage, semi-parametric models of production processes*, Journal of Econometrics 136 (2007), no. 1, 31–64. https://doi.org/10.1016/j.jeconom.2005.07.009.

32. K. Boakye et al., *Data envelopment analysis to estimate technical and scale efficiencies of smallholder pineapple farmers in Ghana*, Agriculture 14 (2024), no. 7, 1032. https://doi.org/10.3390/agriculture14071032.

33. A. Santosa et al., *Technical efficiency of sugarcane farming in East Java, Indonesia: A bootstrap data envelopment analysis*, Open Agriculture 9 (2024), no. 1, 20220378. https://doi.org/10.1515/opag-2022-0378.

34. R. Mulwa, A. Emrouznejad and L. Muhammad, *Economic efficiency of smallholder maize producers in Western Kenya: A DEA meta-frontier analysis*, International Journal of Operational Research 4 (2009), no. 3, 250–267. https://doi.org/10.1504/IJOR.2009.023284.

35. H. M. A. Ilyas et al., *Energy efficiency outlook of New Zealand dairy farming systems: An application of data envelopment analysis*, Energies 13 (2020), no. 1, 251. https://doi.org/10.3390/en13010251.

36. S. Abdulai et al., *Assessing the technical efficiency of maize production in northern Ghana: The data envelopment analysis approach*, Cogent Food and Agriculture 4 (2018), no. 1, 1512390. https://doi.org/10.1080/23311932.2018.1512390.

37. S. Lozano et al., *Forest efficiency assessment and prediction using dynamic DEA and machine learning*, Forest Policy and Economics 154 (2025), 103461. https://doi.org/10.1016/j.forpol.2025.103461.

38. S. Ajith et al., *Yield prediction, pest and disease diagnosis, soil fertility mapping, precision irrigation scheduling, and food quality assessment using machine learning and deep learning algorithms*, Discover Food 5 (2025), no. 1, 1–23. https://doi.org/10.1007/s44187-025-00338-1.

39. N. M. Guerrero, J. Aparicio and D. Valero-Carreras, *Combining data envelopment analysis and machine learning*, Mathematics 10 (2022), no. 6, 909. https://doi.org/10.3390/math10060909.

40. L. Chen et al., *A hybrid data envelopment analysis–random forest methodology for evaluating green innovation efficiency in an asymmetric environment*, Symmetry 16 (2024), no. 8, 960. https://doi.org/10.3390/sym16080960.

41. Z. Zhang, Y. Xiao and H. Niu, *DEA and machine learning for performance prediction*, Mathematics 10 (2022), no. 10, 1776. https://doi.org/10.3390/math10101776.

42. S. Boubaker et al., *Predicting the performance of MSMEs: A hybrid DEA–machine learning approach*, Annals of Operations Research 350 (2025), no. 2, 555–577. https://doi.org/10.1007/s10479-023-05230-8.

43. G. Cattani, *Combining data envelopment analysis and random forest for selecting optimal locations of solar PV plants*, Energy and AI 11 (2022), 100222. https://doi.org/10.1016/j.egyai.2022.100222.

44. N. Zhu, C. Zhu and A. Emrouznejad, *A combined machine learning algorithms and DEA method for measuring and predicting the efficiency of Chinese manufacturing listed companies*, Journal of Management Science and Engineering 6 (2021), no. 4, 435–448. https://doi.org/10.1016/j.jmse.2020.10.001.

A. Srikrishna,
Research Scholar,
Department of Mathematics,
Vellore Institute of Technology,
Chennai,
Tamil Nadu,
India
First Author.
E-mail address: `srikrishna.a2023@vitstudent.ac.in`

and

G. Y. Mythili,
Assistant Professor Senior Grade 1,
Department of Mathematics,
Vellore Institute of Technology,
Chennai,
Tamil Nadu,
India
Corresponding Author.
E-mail address: `mythili.gy@vit.ac.in`