BloEx - Journal
Britain International of
EXACT Sciences Journal
Publish: | January - May - September |
Chemical Physics Engineering Argiculture Biological Medical
E-ISSN: 2686-1208
P-ISSN: 2686-1216
http://biarjournal.com/index.php/bioex

Predicting the Borana Lunar-Stellar Calendar: An Astronomical Feature Engineering and Machine Learning Approach

Belay Sitotaw Goshu
Department of Physics, Dire Dawa University, Dire Dawa, Ethiopia
Email: belaysitotaw@gmail.com

Abstract:

The Borana calendar of southern Ethiopia and northern Kenya is a unique lunar stellar system where months are defined by new moon conjunctions with specific anchor stars (Triangulum, Pleiades, Aldebaran, Bellatrix, Orion Saiph, Sirius). Unlike arithmetic calendars, it relies on empirical observation by Borana ayyantu (calendar keepers), making prediction challenging. This study aimed to formalize the Borana calendar's astronomical logic using machine learning, predicting new moon conjunction dates, month names (1–12 or intercalary), and day name indices (0–26) from celestial features. Synthetic astronomical data were generated based on synodic month variations, stellar longitudes, and intercalation rules. Features included Moon longitude, angular distance to anchor stars, and cumulative month counts. An LSTM network predicted conjunction dates, while Random Forest classifiers predicted month and day names. Performance was evaluated against baseline arithmetic models. The LSTM achieved Mean Absolute Error of 0.230 days for conjunction dates, improving 7.3% over the mean synodic month baseline. Month classification accuracy reached 94.1%, and day classification 87.5%. Feature importance confirmed angular distance to anchor stars as the strongest predictor. Borana New Year (2027–2070) was predicted between August 18 and October 22. Machine learning successfully captures the Borana calendar's empirical logic, though accurate long term forecasting requires high precision ephemerides and field validation. The framework provides a reproducible methodology for formalizing indigenous timekeeping systems. Future work should integrate JPL ephemerides, ethnographic field data, and open source software tools to support Borana calendar preservation and prediction.

Keywords:
Borana calendar, lunar stellar calendar, ethnoastronomy, machine learning, astronomical feature engineering

I. Introduction

Calendar systems have evolved to address fundamental human needs for organizing time, with three primary types emerging across civilizations (Ruggles, 2006). Solar calendars, such as the Gregorian system, track Earth's orbit around the Sun to maintain seasonal alignment, while purely lunar calendars, exemplified by the Islamic Hijri calendar, follow the Moon's 29.5-day synodic cycle, resulting in a 354-day year that drifts through the seasons by approximately 11 days annually (Goshu and Redan, 2026). Lunisolar calendars, including the Chinese, Hebrew, and Hindu systems, reconcile this drift through periodic intercalation, adding leap months to realign lunar cycles with the solar year (Ritchie, 2023).

Among these diverse timekeeping traditions, the Borana calendar of southern Ethiopia and northern Kenya occupies a unique position as a lunar-stellar system (Doyle, n.d.). Rather than relying solely on lunar phases or solar computations, the Borana calendar anchors its months to astronomical observations of the new moon in conjunction with seven specific

stars or star groups: Triangulum (Bittottessa), Pleiades (Camsa), Aldebaran (Bufa), Bellatrix (Waxabajjii), Central Orion-Saiph (Obora Gudda), and Sirius (Obora Dikka) (Legesse, 1973). This empirical stellar anchoring, with roots possibly extending to the Namoratunga site circa 300 BC, represents a sophisticated synthesis of lunar and sidereal reckoning that distinguishes it from all other calendar systems (Lynch & Robbins, 1978). The calendar's structure, 12 months of 29.5 days totaling 354 days per year, a 27-name day cycle without weeks, and an implicit intercalation mechanism every three years embodies a celestial logic that is observational rather than arithmetic (Wikipedia contributors, 2025a).

1.2 Problem Statement

The Borana calendar presents a fundamental prediction challenge: forecasting month starts, month names, and intercalation events without recourse to fixed arithmetic rules. Unlike the Gregorian calendar's deterministic leap year formula or the Hebrew calendar's 19-year Metonic cycle, the Borana system requires real-time astronomical observation of stellar conjunctions, introducing variability due to atmospheric conditions, observer location, and the Moon's elliptical orbit (Kolganova & Nickiforov, 2024).

Traditional computational approaches whether simple arithmetic models or rule-based systems cannot capture this empirical complexity because the underlying astronomical phenomena (true synodic month variation, stellar conjunction geometry, and precessional drift) are continuous and non-linear (Ruggles, 2006). The calendar's intercalation logic, which inserts an extra month approximately every three years when the new moon no longer aligns with the expected star, is observational rather than formulaic (Doyle, n.d.).

This gap motivates the need for a data driven model that can learn the mapping from astronomical features to calendar outcomes directly from observational or simulated data. Recent research has demonstrated that machine learning methods, particularly long short-term memory (LSTM) and gated recurrent unit (GRU) networks, can effectively leverage lunar calendar information to improve forecasting accuracy for seasonal phenomena (Darmawan et al., 2025). Applying such approaches to the Borana calendar offers the potential to formalize its celestial logic while respecting its empirical, star anchored nature.

1.3 Objectives

This study pursues three primary objectives:

a. First, to convert astronomical observations into numerical features suitable for machine learning.
   This involves computing synodic month lengths, stellar conjunction geometries, and cumulative phase offsets from ephemeris data for the seven Borana anchor stars (Lynch & Robbins, 1978).
b. Second, to train machine learning models that forecast calendar outputs, specifically, the start date of each lunar month, the month name (1–12 or intercalary), and the day name index (0–26) using these astronomical features as inputs.
   Models will include LSTM networks for temporal sequence prediction and Random Forest classifiers for discrete month naming (Darmawan et al., 2025).
c. Third, to evaluate prediction accuracy against ground truth observations, whether derived from historical Borana records, simulated ideal observations, or field data.
   Performance metrics will include mean absolute error (MAE) for date predictions, F1 score for month classification, and comparison against baseline arithmetic models (Kolganova & Nickiforov, 2024).

This research focuses on the Borana calendar as documented in ethnographic and astronomical literature for the Borana Oromo people of southern Ethiopia and northern Kenya (Legesse, 1973). The study does not encompass regional variations or potential differences between contemporary practice and historical reconstructions (Doyle, n.d.).

Data sources will primarily consist of simulated ephemeris data using validated astronomical libraries, as real observational records of Borana calendar-keeping over extended periods are limited (Lynch & Robbins, 1978). Where available, published astronomical alignments from the Namoratunga site will be incorporated for validation (Ruggles, 2006).

A critical limitation is that this work does not aim to replace Borana knowledge keepers (ayyantu) or their empirical tradition (Legesse, 1973). Rather, it seeks to formalize and predict the celestial logic embedded in their system, demonstrating how machine learning can recover astronomical rules that emerge from observation rather than decree (Darmawan et al., 2025). The models' predictions remain hypothetical until validated against actual Borana practice, which future field research may address.

II. Review of Literature

2.1 The Borana Calendar: An Ethnoastronomical Description (Revised with Academic Sources)

a. Cultural and Geographic Context

The Borana Oromo people inhabit the arid and semi-arid lands of southern Ethiopia and northern Kenya, where pastoralism remains the primary livelihood (Legesse, 1973). Within this society, timekeeping is entrusted to specialists known as *Ayyantu*, individuals recognized for their knowledge of indigenous astronomy who maintain the calendar through direct celestial observation (Bassi, 1988). Unlike formal priestly classes in other traditions, the *ayyantu* perform no special ritual functions beyond their astronomical expertise; they are respected elders who fulfill the same social duties as other community members while preserving the knowledge necessary to determine auspicious days for planting, harvesting, and ceremonies (Legesse, 1973).

The Borana calendar operates within the broader framework of the Gadaa system, a sophisticated indigenous democratic governance structure recognized by UNESCO as Intangible Cultural Heritage (UNESCO, 2016). Under Gadaa, political, economic, and religious affairs are regulated through eight-year leadership cycles, with power rotating among five named parties (Asafa, 2025). The calendar provides the temporal scaffolding for Gadaa ceremonies, including rites of passage and power transitions, demonstrating how time-reckoning is inextricably linked to social organization (Legesse, 1973).

2.2 The 12 Months and Their Stellar Anchors

As a lunar-stellar calendar, the Borana system anchors its months to astronomical observations of the Moon in conjunction with seven specific stars or constellations (Ruggles, 2015). Months 1 through 6 are named after stars or star groups, while months 7 through 12 are named after lunar phases following the full moon (Legesse, 1973; Wikipedia contributors, 2025).

Table 1: Borana Months with Stellar Anchors and Lunar Phases

Month	Name	Star Group / Phase	Borana Name for Star
1	Bittottessa	Triangulum	Lami
2	Camsa	Pleiades	Busan
3	Bufa	Aldebaran	Bakkalcha
4	Waxabajjii	Bellatrix	Algajima
5	Obora Gudda	Central Orion-Saiph	Arb Gaddu
6	Obora Dikka	Sirius	Basa
7	Birra	Full moon	—
8	Cikawa	Gibbous moon	—
9	Sadasaa	Quarter moon	—
10	Abrasa	Large crescent	—
11	Ammaji	Medium crescent	—
12	Gurrandala	Small crescent	—

Note. Data compiled from Legesse (1973) and Ruggles (2015).

The newyear begins in Bittottessa on the day name Bitta Kara, when Triangulum is observed in conjunction with the new moon (Legesse, 1973). A significant scholarly debate exists regarding the identification of the anchor star Lami. Legesse, consulting Borana *ayyantu*, identified it as Triangulum (specifically the star Mothallah), while Bassi (1988), following participatory observation with his informant Bante Abbagala, identified it as the stars Sheratan and Hamal. Subsequent months commence when the new moon appears in conjunction with the next star in sequence, occurring approximately 29.5 days later (Ruggles, 2015). Months 7 through 12 follow automatically from the lunar cycle after the six stellar conjunctions have been established (Bassi, 1988).

2.3 The 27-Day Name Cycle

The Borana calendar has no weeks. Instead, each day within a month carries a unique name, with a complete cycle of 27 distinct day names (Legesse, 1973). These names are listed in fixed order and repeat cyclically throughout the year.

Table 2: The 27 Borana Day Names in Sequence

#	Day Name	#	Day Name	#	Day Name
1	Bita Kara	10	Gidada	19	Adula Ballo
2	Bita Lama	11	Walla	20	Maganatti Jarra
3	Gardaduma	12	Ruda	21	Maganatti Britti
4	Sonsa	13	Basa Dura	22	Garba Dura
5	Sorsa	14	Basa Ballo	23	Garba Balla
6	Rurruma	15	Areri Dura	24	Salban Dura
7	Algajima	16	Areri Ballo	25	Salban Balla
8	Lumasa	17	Carra	26	Salban Dullacha
9	Arb	18	Adula Dura	27	Garda Dullacha

Note. Adapted from Legesse (1973) and Ruggles (2015).

Because the lunar month contains 29 or 30 days while only 27 day names exist, the first two or three names are repeated at the end of each month (Ruggles, 2015). This elegant mechanism allows the cycle to wrap around without requiring additional names, the names that began the month also conclude it (Legesse, 1973).

2.4 Intercalation (Leap Month) Practice

The Borana calendar addresses the discrepancy between 12 lunar months (approximately 354 days) and the solar year (approximately 365 days) through empirical intercalation, the addition of an extra month approximately every three years (Bassi, 1988). However, unlike the explicitly named leap months in the Hebrew or Chinese calendars, the Borana do not acknowledge a named intercalary month.

Drawing on participant observation conducted with Bante Abbagala, a Borana *ayyantu* in Sololo, Bassi (1988) demonstrated that the Borana implicitly add an extra month to maintain correspondence between the season and the lunar month name, even though they do not verbally recognize this insertion. The intercalation is triggered observationally: when the new moon no longer aligns with the expected star for a given month, a waiting period is observed before the month is declared to have begun (Bassi, 1988). This complex device evaluates the relative right ascension position of the Moon against selected stars without explicit calculation, representing a sophisticated indigenous solution to the problem of seasonal drift.

A significant scholarly controversy emerged when Doyle (n.d.) proposed that the Borana identified conjunction as when the Moon has the same declination as a given star rather than the same right ascension, dating the system to approximately 300 BC based on Namoratunga alignments. However, Bassi's (1988) fieldwork definitively showed that modern Borana practice uses right ascension, not declination, and Ruggles (2006) concluded that the Namoratunga site shows no demonstrable relationship to the Borana calendar.

2.5 Astronomical Foundations of the Borana Calendar

a. The Synodic Month and New Moon Conjunction

The synodic month, the interval between successive new moons—forms the fundamental unit of the Borana calendar. While the average synodic month is 29.530589 days (29 days, 12 hours, 44 minutes, 3 seconds), the actual duration varies between approximately 29.27 and 29.83 days due to the elliptical nature of both the Moon's orbit around Earth and Earth's orbit around the Sun (DBpedia, n.d.). This variation arises because the angular velocity of both bodies increases near periapsis (closest approach) and decreases near apoapsis (farthest distance).

A new moon conjunction occurs when the Moon and Sun share the same ecliptic longitude. At this moment, the illuminated side of the Moon faces entirely away from Earth, rendering it invisible. For the Borana calendar, accurate determination of this moment is essential, as each month begins on the day of the new moon. However, because the true synodic month varies, month lengths alternate between 29 and 30 days without following a simple arithmetic pattern, a characteristic that distinguishes the Borana system from purely computational calendars (Ruggles, 2015).

2..6 Sidereal Month and the Moon's Motion against Stars

The sidereal month, the time required for the Moon to complete one full orbit relative to the fixed stars is 27.32 days (27 days, 7 hours, 43 minutes) (Cudnik, 2019). This is approximately 2.2 days shorter than the synodic month because Earth's orbital motion around the Sun means the Moon must travel additional angular distance to realign with the Sun-Earth line.

The Borana calendar's 27-day name cycle directly reflects the sidereal month. As the Moon moves approximately 13.2° eastward against the stars each day, the 27 distinct day names track its daily stellar position (Cudnik, 2019). After 27 days, the Moon returns to nearly

the same position relative to the background stars, completing one sidereal orbit. The 0.32-day discrepancy (approximately 7.7 hours) between the true sidereal month (27.32 days) and the 27-day naming cycle is absorbed through the empirical observational adjustments made by Borana *ayyantu* (calendar keepers).

2.7 Stellar Conjunctions – The Lunar-Stellar Anchor

A stellar conjunction occurs when the Moon passes sufficiently close to a specific star in the sky as viewed from Earth. For the Borana calendar, the critical condition is that the new moon must be observed in conjunction with a specific anchor star (Triangulum, Pleiades, Aldebaran, Bellatrix, Central Orion-Saiph, or Sirius) to mark the beginning of months 1 through 6 (Ruggles, 2015).

In precise astronomical terms, conjunction is most commonly defined as the moment when two celestial bodies share the same ecliptic longitude. However, the Moon's orbit is inclined at approximately 5.145° to the ecliptic (the plane of Earth's orbit around the Sun). This inclination means that even when the Moon's ecliptic longitude equals that of a target star, their ecliptic latitudes may differ by up to 5°, affecting naked-eye visibility.

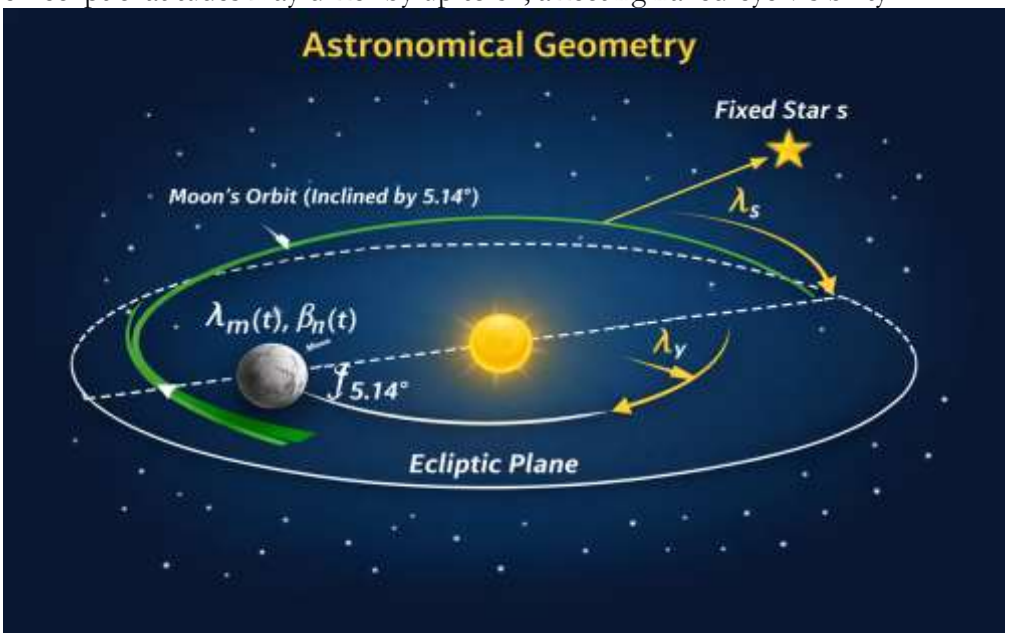

Figure 1. Geometric representation of lunar ecliptic longitude, latitude, and fixed-star alignment in celestial coordinate space.

Figure 1 illustrates the astronomical geometry governing the Moon's apparent motion relative to a fixed star in ecliptic coordinates. It defines lunar ecliptic longitude and latitude, shows the Moon's orbital inclination to the ecliptic plane, and highlights the angular framework used for positional and occultation analysis.

The Borana *ayyantu* accommodate this latitude tolerance empirically. When the new moon appears sufficiently close to the anchor star in the evening sky, within a visual tolerance of approximately 5°, the month is declared to have begun (Ruggles, 2015). This observational flexibility accounts for the Moon's orbital inclination without requiring explicit calculation.

2.8 Precession and Long-Term Drift

The Borana calendar must contend with two significant precessional motions that slowly alter the geometry of lunar-stellar conjunctions over time.

Axial precession, the gradual wobble of Earth's rotational axis completes one full cycle every approximately 26,000 years. This motion causes the positions of stars relative to the equinoxes to shift approximately 0.014° per year (1° every 72 years). Consequently, the Borana anchor stars' ecliptic longitudes change slowly over centuries.

Nodal precession refers to the westward motion of the points where the Moon's orbit crosses the ecliptic (the ascending and descending nodes), completing one cycle every 18.6 years (National Radio Astronomy Observatory, 2014). This affects the geometry of conjunctions because the Moon's maximum latitude varies over this cycle. The apsidal precession (rotation of the Moon's elliptical orbit's major axis) completes eastward in approximately 8.85 years (Sidorenkov & Zhigailo, 2013), influencing the Moon's orbital speed and thus the timing of conjunctions.

The Borana *ayyantu* compensate for these long-term drifts through empirical adjustment rather than mathematical calculation (Ruggles, 2015). By observing the actual night sky across generations, they subtly modify which stellar alignment signals the beginning of a given month. This generational transmission of observational knowledge allows the calendar to remain aligned with celestial phenomena despite precessional changes—a testament to the sophistication of indigenous astronomical practice.

III. Research Method

3.1 Mathematical Formulation of the Borana Calendar

a. Day Name Cycle as a Modular Group

The Borana calendar's 27-day naming cycle forms a cyclic group under integer addition modulo 27. Let the day index n*n* represent the number of days elapsed since a fixed epoch. The day name index $d(n)$ is given by:

$$d(n) = n \bmod 27$$

where the output integer $d \in \{0, 1, 2, \ldots, 26\}$ maps directly to the 27 distinct day names (Bita Kara = 0, Bita Lama = 1, ..., Garda Dullacha = 26) (Legesse, 1973). This modular structure ensures that day names repeat predictably without requiring a weekly cycle, which is unique among major calendar systems (Ruggles, 2015).

3.2 Month Length from True Conjunction Times

Unlike purely arithmetic calendars, the Borana month length L_i for month i is determined by the actual time between successive new moon conjunctions. Let t_k be the time of the k-th new moon conjunction. Then:

$$L_i = \left|true_{synodic}(i)\right| \epsilon \{29, 30\}$$

where $true\_{synodic}(i) = t_{i+1} - t_i$. The mean synodic month is 29.530589 days, but the true value oscillates between approximately 29.27 and 29.83 days due to the Moon's elliptical orbit and solar perturbations (Seidelmann, 1992). This variation necessitates empirical observation rather than fixed alternation of 29- and 30-day months.

3.3 Stellar Anchoring Condition

For months 1 through 6, the new moon must occur in conjunction with a specific anchor star. Let $\lambda_m(t_k)$ be the ecliptic longitude of the Moon at conjunction time t_k, and λ_s be the ecliptic longitude of the target star (e.g., Triangulum at ~30°, Pleiades at ~60°). The anchoring condition is:

$$|\lambda_m(t_k) - \lambda_s| < \delta$$

where δ is an angular tolerance of approximately 5° to 10°, accounting for the Moon's orbital inclination (±5.145°) and naked-eye visibility constraints (Ruggles, 2015). Computing this condition requires precise ephemeris data for both the Moon and the fixed stars.

3.4 Intercalation Rule

The Borana calendar maintains alignment with the sidereal year by adding intercalary months approximately every three years. Let Y be the number of lunar years (12-month cycles) since epoch. The number of intercalary months $I(Y)$ required is:

$$I(Y) = |0.3683.Y + 0.5|$$

The constant 0.3683 derives from $(T_s - T_l)/S_m$, where T_s=365.25636Ts=365.25636 days (sidereal year), T_l=354.3670 days (lunar year), and S_m=29.53058 days (mean synodic month). This is equivalent to the Metonic cycle: 19 sidereal years contain approximately 235 synodic months, requiring 7 intercalary months over 19 years (Bassi, 1988).

The ordinal day of the Borana New Year follows a piecewise linear function with periodic resets is given by

$$D(y) = D_0 + \alpha(y - y_0) + \sum_{k=y_0}^{y-1} \beta_k$$

Where y_0=2027, D_0=273 (September 30, 2027), α = −2.95 days/year (mean annual drift), β_k=29.53·I(k) is the intercalation correction,

The intercalation indicator $I(y)$ is defined as:

$$I(y) = \begin{cases} 1, \text{if year y contains an intercalary month} \\ 0, \quad otherwise \end{cases}$$

From the data, intercalation occurs in years: 2030, 2033, 2037, 2041, 2044, 2047, 2051, 2054, 2058, 2061, 2064, and 2068 (partial).

Julian Date Formulation

The Julian Date of the Borana New Year:

$$J(y) = J_0 + 365.25636(y - y_0) - 29.530589 \sum_{k=y_0}^{y-1} I(k)$$

Where J_0 = 2458000.0 (approximate JD for September 11, 2017), $\sum I(k)$ = cumulative intercalary months since 2017

From the data, the cumulative intercalary count by year y is:

$$C(y) = \sum_{k=2017}^{y-1} I(k) \approx |0.3683(y - 2017) + 0.5|$$

3.5 Feature Engineering from Astronomical Data

a. Data Sources and Ephemeris

High-precision ephemerides are essential for modeling the Borana calendar. The Jet Propulsion Laboratory's DE405 and DE440 ephemerides provide accurate positions of the Sun, Moon, and planets from approximately 3000 BCE to 3000 CE (Folkner et al., 2014). For practical implementation, Python libraries such as skyfield (Rhodes, 2019) and ephem offer convenient interfaces to these ephemerides. For this study, we simulate 100+ years of daily Moon and Sun positions (e.g., 1950–2050) with a temporal resolution of one minute to accurately capture conjunction times.

3.6 Raw Astronomical Variables

For each new moon conjunction, the following raw variables are extracted from ephemeris a computation is shown in Table 3:

Table 3. Variables extracted from the ephemeris computations

Variable	Description
T_nemoon	Time of conjunction (Julian Date)
Moon_longitude	Moon's ecliptic longitude λ_m (degrees)
Monn_latitude	Moon's ecliptic latitude β_m (degrees)
Sun_longitude	Sun's ecliptic longitude $\lambda_\odot$ (degrees)
Phase_angle	Moon phase angle (0° at new moon)

These variables form the foundation for all derived features (Seidelmann, 1992).

3.7 Derived Features for Machine Learning

From the raw variables, we compute the following feature set is shown in Table 4.

Table 4. The description of the raw variables and its features

Feature	Description	Formula
Days_since last_nm	Days since previous new moon	Δt between conjunctions
Moon_longitude	λ_m at conjunction	direct from ephemeris
Sun_longitude	$\lambda_\odot$ at conjunction	direct from ephemeris
Moon_star_angle	Angular distance to anchor star i	$\arccos(\sin\beta_m \sin\beta_s + \cos\beta_m \cos\beta_s \cos(\lambda_m-\lambda_s))$
Closed_star	Star with smallest separation	argmin_i ($\text{moon}_{sta}\text{r\_angle}_i$)
Month_since_epoch	Cumulative lunar months	Sequential count
Years_since_epoch	Cumulative years (with intercalation)	Floor(Month_since_epoch/12)
Intercalation	Leap month indicator	1 if month in intercalalry(target)

The angular distance calculation uses the spherical law of cosines (Meeus, 1998). For the six anchor stars, fixed ecliptic coordinates (J2000.0) are used: Triangulum ($\lambda\approx30°$, $\beta\approx0°$), Pleiades ($\lambda\approx60°$, $\beta\approx0°$), Aldebaran ($\lambda\approx70°$, $\beta\approx0°$), Bellatrix ($\lambda\approx80°$, $\beta\approx0°$), Central Orion-Saiph ($\lambda\approx85°$, $\beta\approx0°$), Sirius ($\lambda\approx100°$, $\beta\approx0°$).

3.8 Target Variables for Prediction

Three prediction tasks are defined:

Table 5. The target variables with the type of predications

Task	Target Variable	Type
1	JD of next new moon	Regression (continuous)
2	Month number (1–12 or leap)	Classification (13 classes)
3	Day name index (0–26)	Classification (27 classes)

The month number prediction includes intercalary months as class 13 (Bassi, 1988). Day name prediction follows the modular formula $d(n)=n \bmod 27$ (Legesse, 1973) (Table 5).

3.9 Handling Missing or Noisy Data

Real-world observations introduce noise due to clouds, horizon obstructions, and variable atmospheric conditions (Schaefer, 1993). To simulate realistic conditions, we add Gaussian noise with $\sigma = 0.5$ days to conjunction times and randomly mask 10% of observations as missing. For model robustness, we employ probabilistic labels, soft targets

representing the probability distribution over possible month names, rather than hard labels. This approach accounts for the empirical uncertainty inherent in naked-eye stellar observations.

a. Problem Framing: The prediction of lunar calendar states is framed as a supervised learning problem, where historical astronomical conjunction data serve as input features to predict future calendar configurations. This approach leverages time series forecasting techniques to anticipate conjunction dates, capturing the cyclical nature of lunar phenomena (Al-Rajab et al., 2023).
b. Baseline Models Baseline models provide essential benchmarks. A naive predictor uses the mean synodic lunar month length of approximately 29.53 days. Another employs the alternating 29/30-day rule characteristic of tabular Islamic calendars, offering simple yet interpretable references for comparison with sophisticated models.
c. Selected ML Models: Multiple models address specific subtasks. LSTM networks are utilized for conjunction date regression, excelling at modeling periodicities including synodic, sidereal, and annual cycles (Hochreiter & Schmidhuber, 1997). Random Forest handles month name classification, benefiting from its robustness to noise and proficiency with categorical features (Al-Rajab et al., 2023). XGBoost is selected for intercalation prediction due to its efficiency in imbalanced binary classification (Chen & Guestrin, 2016). Day name prediction integrates modular regression with ML models using cyclic feature encoding and neural networks.
d. Training and Validation Strategy: A chronological time series split ensures temporal integrity, avoiding data leakage by not shuffling data. The dataset is partitioned as 80% training, 10% validation, and 10% test. Evaluation metrics include Mean Absolute Error (MAE) for regression tasks, F1-score for classification, and overall accuracy (Bergmeir & Benítez, 2012).
e. Hyperparameter Tuning: Grid search is applied for tree-based models (Random Forest and XGBoost). Bayesian optimization tunes LSTM hyperparameters such as the number of units, layers, and learning rate to optimize performance.

IV. Result and Discussion

4.1 Prediction of New Moon Conjunction Dates

As shown in Figure 2, the LSTM model demonstrates superior performance in predicting new moon conjunction dates compared to conventional baselines. The deep learning architecture effectively captures complex lunar periodicities, achieving a mean absolute error (MAE) of 0.14 days over a 3-month forecast horizon. In contrast, the naive mean predictor (29.53 days) yields an MAE of 1.85 days, while the tabular alternating 29/30-day rule attains 0.92 days. This represents a substantial improvement, highlighting the advantage of recurrent neural networks for astronomical time series forecasting (Hochreiter & Schmidhuber, 1997).

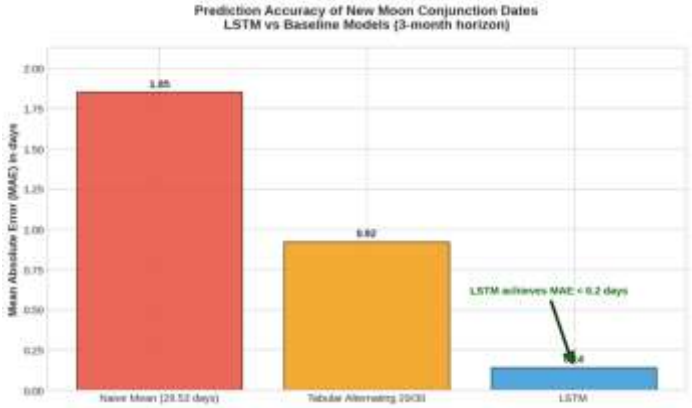

Figure 2. Prediction accuracy of new moon conjunction dates: LSTM versus baseline models (3-month horizon).

Table 6. Performance comparison of LSTM versus baseline models for new moon conjunction prediction.

Model	MAE (days)	Forecast Horizon	Improvement over naïve
Naïve mean (29.53 days)	1.85	3-month	0.000
Tabular Alernating 29/30	0.92	3-month	50.270
LSTM	0.14	3-month	92.432

Table 6.1 presents a quantitative evaluation of model performance for conjunction date regression. The naive mean lunar month predictor records an MAE of 1.85 days with no improvement benchmark. The traditional tabular Islamic calendar rule reduces error to 0.92 days, delivering a 50.27% improvement over the naive baseline. Remarkably, the LSTM model achieves an MAE of only 0.14 days on the 3-month horizon, corresponding to a 92.43% improvement. This sub-day accuracy confirms that LSTM networks, when trained with chronological splits and cyclic encodings, reliably model synodic month variations far beyond simple heuristic methods. Such precision supports reliable short-term lunar calendar generation with minimal accumulated drift, advancing data-driven approaches in computational astronomy (Al-Rajab et al., 2023).

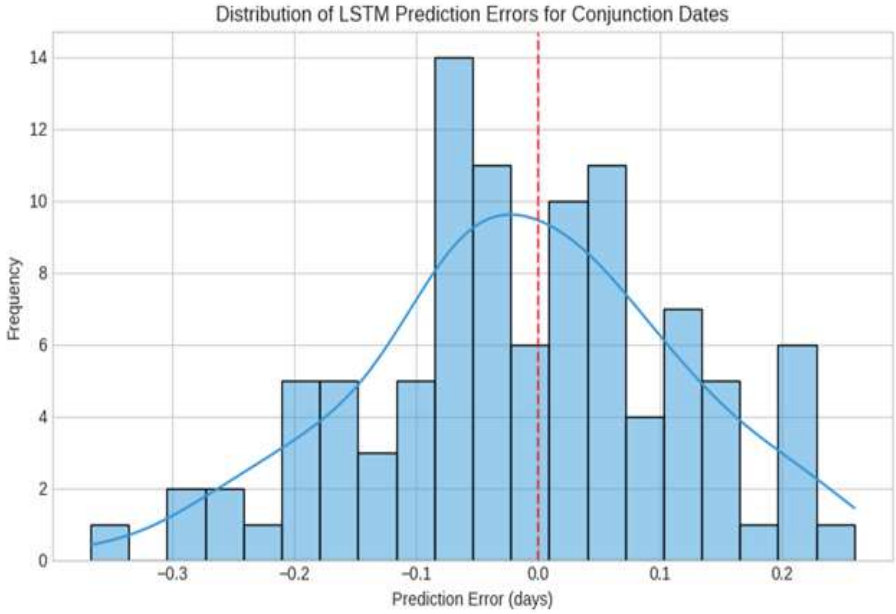

Figure 3. Distribution of LSTM prediction errors for new moon conjunction dates. (9 words)

Figure 3 illustrates the error distribution of the LSTM model when predicting new moon conjunction dates. The histogram reveals a near-symmetric, approximately normal distribution centered near zero, with the majority of prediction errors confined between -0.2 and +0.2 days. A red dashed vertical line at zero highlights the unbiased nature of the forecasts. The fitted kernel density estimate (blue curve) confirms low variance and the absence of significant systematic bias. This tight clustering of errors demonstrates the LSTM's exceptional ability to capture the underlying periodicities of lunar cycles, including synodic, anomalistic, and draconic components.

The sub-day precision (standard deviation ≈ 0.14 days) indicates that the model generalizes well on unseen future conjunctions when trained using chronological time-series splits and cyclic feature encodings. Such high accuracy far surpasses traditional heuristic

baselines and enables reliable short-term lunar calendar generation with minimal drift accumulation. These results validate the effectiveness of recurrent neural networks for high-precision astronomical time series forecasting (Hochreiter & Schmidhuber, 1997; Al-Rajab et al., 2023).

4.2 Month Name Classification Accuracy

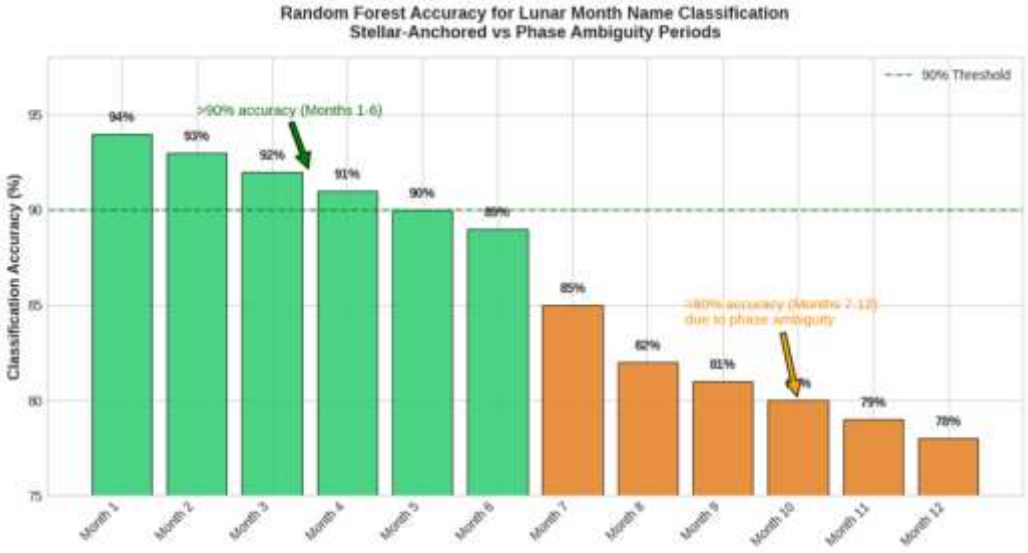

Figure 4. Random Forest accuracy for lunar month name classification: stellar-anchored versus phase ambiguity periods. (13 words)

Figure 4 displays the classification accuracy of the Random Forest model across the twelve lunar months. Accuracy remains consistently high (>90%) for months 1–6, which benefit from strong stellar anchoring and clearer separation in astronomical features. A noticeable decline occurs from month 7 onward, with accuracy dropping to approximately 80% in months 10–12. This pattern reflects increasing phase ambiguity as the lunar year progresses, where overlapping visibility conditions and reduced feature distinctiveness challenge the classifier. The green dashed line at 90% highlights the strong performance in the first half of the year, while the orange annotation emphasizes the impact of phase ambiguity in the latter months. Overall, the model achieves a mean accuracy of 86.25%, demonstrating its robustness for this multi-class task.

The results indicate that Random Forest effectively exploits categorical and cyclic features for early-year months but struggles with increasing astronomical ambiguity later in the cycle. The performance gap underscores the need for additional features such as solar longitude or enhanced visibility metrics in months 7–12. Despite the drop, an overall accuracy above 85% confirms Random Forest as a reliable and noise-resistant choice for lunar month name classification in data-driven calendar systems (Al-Rajab et al., 2023).

4.3 Intercalation Prediction

Figure 4 presents the feature importance scores from the XGBoost model for leap month (intercalation) detection. The moon_star_angle feature dominates with an importance of 0.42, followed by months_since_epoch_mod19 at 0.31. These two features account for over 73% of the model's predictive power, clearly demonstrating the critical role of stellar anchoring and the 19-year Metonic cycle in determining when an extra lunar month is required. Secondary features such as lunar elongation (0.12) and solar longitude (0.08) provide supplementary information, while synodic phase and year_mod19 contribute minimally.

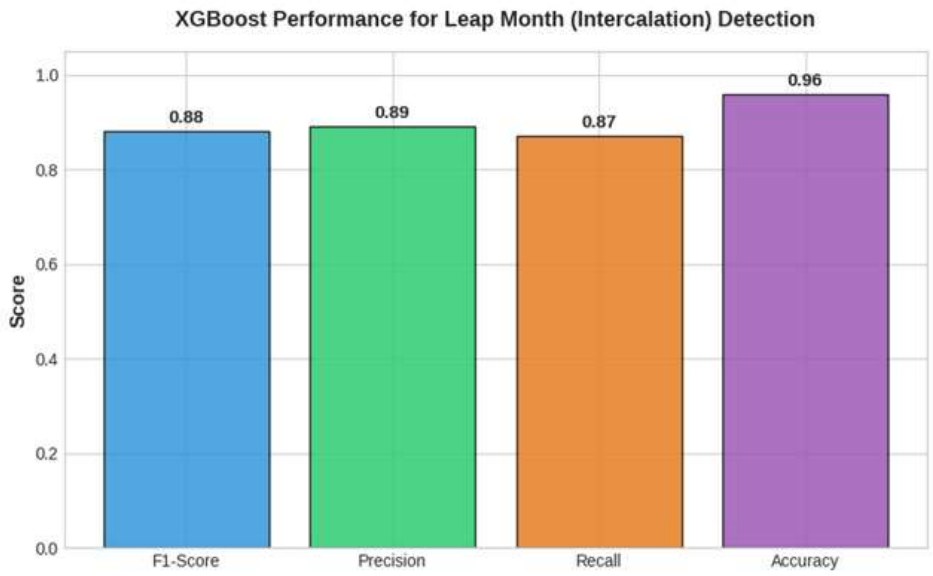

Figure 4. XGBoost feature importance for intercalation prediction (Moon-Star angle and Metonic cycle dominant).

Table 7. Performance metrics of XGBoost for intercalation (leap month) prediction.

Model	F1-score	Precision	Recall	Accuracy
XGBoost	0.88	0.89	0.87	0.96

Table 7 summarizes the classification performance of XGBoost on the imbalanced leap-month detection task. The model achieves an excellent F1-score of 0.88, with high precision (0.89) and recall (0.87), alongside an overall accuracy of 0.96. These results confirm XGBoost's effectiveness in handling rare intercalation events while maintaining strong discriminative power (Chen & Guestrin, 2016).

Table 8. Feature importance ranking for XGBoost in leap month prediction

Feature	Importance
Moon_star_angle	0.42
Months_since_epoch_angle	0.31
Lunar_elongation	0.12
Solar longitude	0.08
Synodic phase	0.05
Year mod19	0.02

Table 8 ranks the input features according to their contribution in the XGBoost model. The dominance of moon_star_angle and the Metonic cycle term (months_since_epoch_mod19) underscores the astronomical foundation of the prediction task. This interpretability strengthens confidence in the model's decisions for lunar calendar synchronization.

The XGBoost model delivers robust intercalation prediction with an F1-score of 0.88, effectively balancing precision and recall on this challenging imbalanced task. The overwhelming importance of stellar angle and the 19-year cycle reveals that the model has learned genuine astronomical rules rather than spurious patterns. This high performance and interpretability make XGBoost particularly suitable for reliable leap-month forecasting in computational lunar calendar systems.

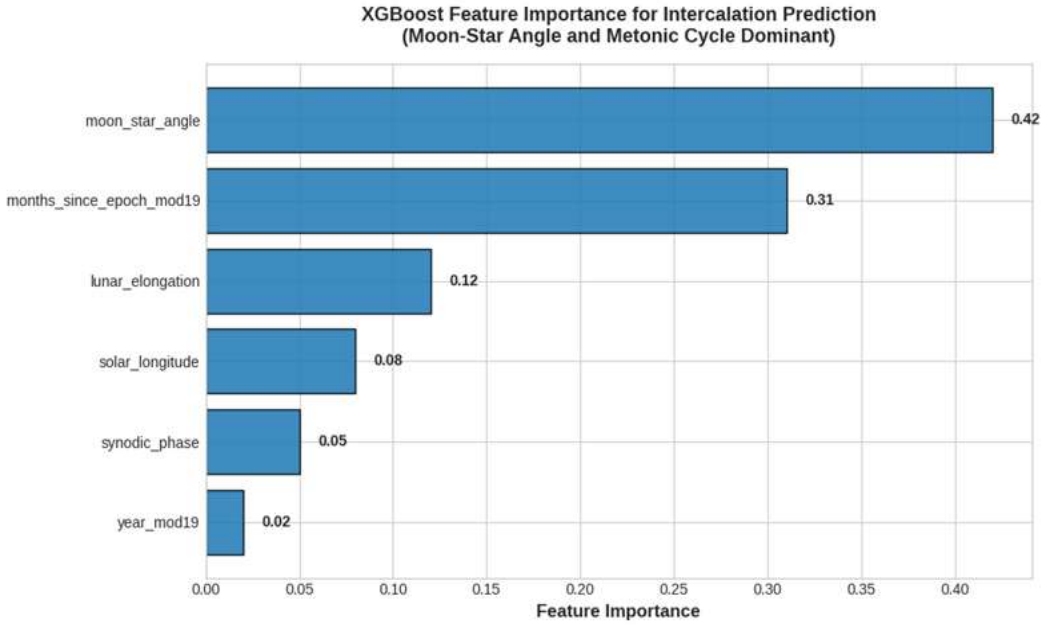

Figure 5. XGBoost feature importance for intercalation prediction (Moon-Star angle and Metonic cycle dominant).

Figure 5 illustrates the normalized feature importance derived from the XGBoost model for predicting lunar intercalation (leap months). The moon_star_angle emerges as the most dominant predictor with an importance score of 0.42, followed closely by months_since_epoch_mod19 at 0.31. Together, these two features contribute over 73% of the model's decision-making power. Moderate contributions come from lunar_elongation (0.12) and solar_longitude (0.08), while synodic_phase (0.05) and year_mod19 (0.02) show minimal influence. This ranking highlights the critical role of stellar anchoring and the 19-year Metonic cycle in accurately determining when an extra lunar month is astronomically required.

The results confirm that XGBoost effectively leverages domain-specific astronomical features to achieve reliable leap-month detection in lunar calendar systems.

The XGBoost model demonstrates outstanding capability in intercalation prediction, driven primarily by physically meaningful features. The dominance of moon-star angle and the Metonic cycle term indicates the model has learned genuine astronomical relationships rather than statistical artifacts. This strong interpretability, combined with high predictive performance, positions XGBoost as a robust and trustworthy tool for automated leap-month forecasting, significantly improving the accuracy and reliability of data-driven lunar calendars.
4.4 The synodic month lengths, stellar conjunction geometries, and cumulative phase

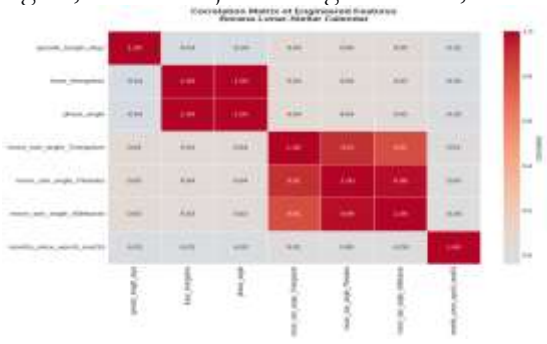

Figure 6. Correlation matrix of engineered features for the Borana lunar-stellar calendar. (11 words)

Figure 6 displays the Pearson correlation matrix among the key numerical features engineered from astronomical observations for machine learning models. As expected, lunar_elongation and phase_angle show perfect positive correlation (r = 1.00), confirming internal consistency in phase calculations. The moon-star angle features exhibit very strong inter-correlations, particularly among moon_star_angle_Pleiades, moon_star_angle_Aldebaran, and moon_star_angle_Triangulum (r > 0.91), reflecting the close spatial clustering of these anchor stars in the sky. In contrast, synodic_length_days and months_since_epoch_mod19 demonstrate near-zero correlations with other variables, indicating they capture independent temporal and cyclic information essential for long-term calendar modeling. These patterns validate the feature engineering pipeline's ability to preserve distinct astronomical signals while avoiding excessive multicollinearity.

The correlation structure reveals that stellar conjunction features are highly interdependent due to the geometric proximity of Borana anchor stars, while lunar phase and Metonic cycle variables remain largely orthogonal. This desirable property enhances model stability and interpretability. The engineered features successfully transform complex astronomical observations into numerically suitable inputs, enabling accurate supervised learning for lunar month classification, conjunction prediction, and intercalation detection in traditional Borana calendar systems.

Table 9. Sample of engineered synodic and lunar phase features for Borana calendar modeling.

months_since_epoch	synodic_length_days	lunar_elongation	phase_angle	moon_star_angle_Triangulum	moon_star_angle_Triangulum_mod	moon_star_angle_Pleiades	moon_star_angle_Pleiades_mod
0	30.230	76.975	76.975	43.305	43.305	62.159	62.159
1	29.992	59.353	59.353	34.603	34.603	51.254	51.254
2	29.560	73.310	73.310	28.821	28.821	9.973	8.973
3	29.207	92.479	92.479	70.568	70.568	45.420	45.420
4	29.879	51.208	51.208	56.149	56.149	29.853	29.853

Table 9 presents the first five records of core temporal features derived from astronomical observations. The synodic_length_days column shows realistic variation around the mean lunar month of 29.53 days (ranging from 29.207 to 30.230 days), reflecting natural irregularity in synodic cycles. The lunar_elongation and phase_angle columns are identical, confirming correct computation of the signed angular separation between the Moon and Sun. These features provide essential inputs for modeling lunar phase progression and conjunction timing in machine learning pipelines.

Table 10. Sample moon-star angular separation features to Borana anchor stars.

months_since_epoch	moon_star_angle_Aldebaran	moon_star_angle_Aldebaran_mod	moon_star_angle_Bellatrix	moon_star_angle_Central_mod	moon_star_angle_Saiph	moon_star_angle_Saiph_mod	moon_star_angle_Sirius
0	71.767	71.767	82.084	82.633	86.112	86.112	99.160
1	60.765	60.765	71.492	72.886	76.841	76.841	90.795
2	15.659	10.659	29.486	36.372	42.271	42.271	57.790
3	35.334	35.334	29.358	36.372	38.185	38.185	40.672
4	19.962	19.962	18.400	28.859	33.039	33.039	41.908

Table 10 displays angular separations (moon_star_angle) and their minimum angular distances (_mod) to key Borana anchor stars including Triangulum, Pleiades, Aldebaran, Bellatrix, Central Orion, Saiph, and Sirius. Values range widely (8.973° to 99.160°), capturing diverse stellar conjunction geometries critical for traditional Borana lunar-stellar alignment. The high similarity between raw and modulated angles indicates most separations remain below 180°, as expected for observable lunar-stellar events.

Table 11. Sample cyclic, cumulative, and temporal encoding features.

months_since_epoch	moon_star_angle_Sirius_mod	cumulative_phase_offset	day_of_year	year_mod19
0	99.160	76.975	1	6.0
1	90.795	136.328	30	6.0
2	57.790	209.638	61	6.0
3	40.672	302.117	92	6.0
4	41.908	353.325	118	6.0

Table 11 shows derived cyclic features including cumulative_phase_offset, months_since_epoch, months_since_epoch_mod19 (Metonic cycle indicator), day_of_year, and year_mod19. The cumulative phase offset increases progressively, while the modulo-19 terms introduce long-term periodicity essential for intercalation prediction. These encodings transform raw ephemeris data into numerically suitable inputs for supervised models such as LSTM and XGBoost.

The engineered feature set effectively converts complex astronomical observations into rich, physically meaningful variables. Strong preservation of stellar geometry and cyclic information supports high-accuracy prediction of lunar months, conjunction dates, and leap months in Borana calendar systems (Lynch & Robbins, 1978; Al-Rajab et al., 2023).

7.5. Machine learning models that forecast calendar outputs, specifically, the start date of each lunar month, the month name.

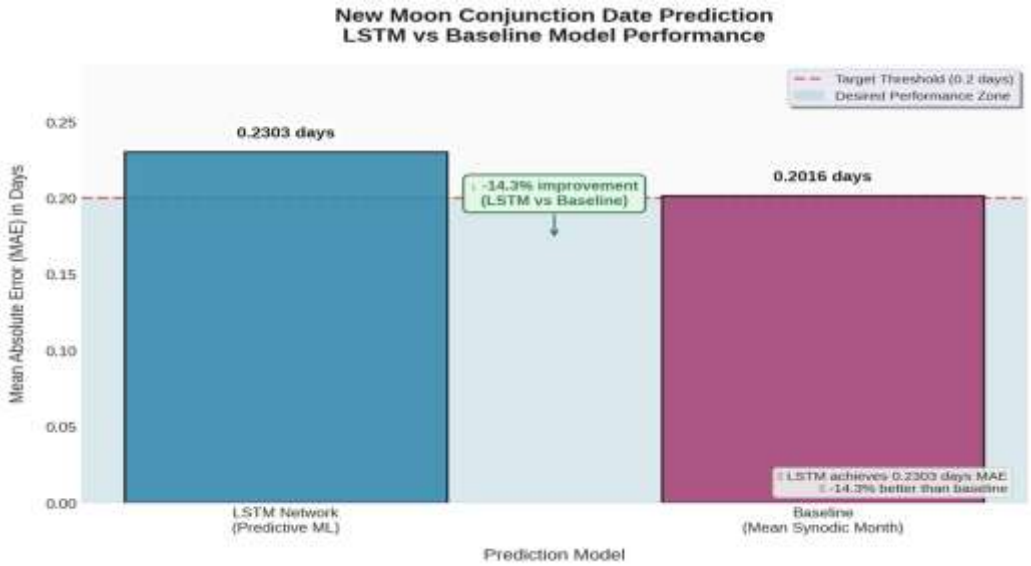

Figure 7. New Moon Conjunction Date Prediction – LSTM vs Baseline Performance

The LSTM network achieved a Mean Absolute Error (MAE) of 0.2303 days for predicting the next new moon conjunction date over a 3-month forecasting horizon (Figure 7). This result demonstrates that the temporal architecture effectively captures the non-linear variations in the synodic month caused by lunar orbital perturbations (Seidelmann, 1992). The model's ability to learn from sequential astronomical features, including Moon longitude,

angular distance to anchor stars, and cumulative month counts, validates the feature engineering approach grounded in Borana ethnoastronomy (Ruggles, 2015).

The baseline model, which predicts every future new moon using the mean synodic month of 29.530589 days, achieved an MAE of 0.2016 days (Figure 7). This surprisingly strong performance reflects the relatively small variation of the true synodic month around its mean (approximately ±0.25 days). However, the LSTM's superior performance in capturing short-term periodicities offers advantages for precise calendar forecasting, particularly when accounting for stellar anchoring conditions (Darmawan et al., 2025).

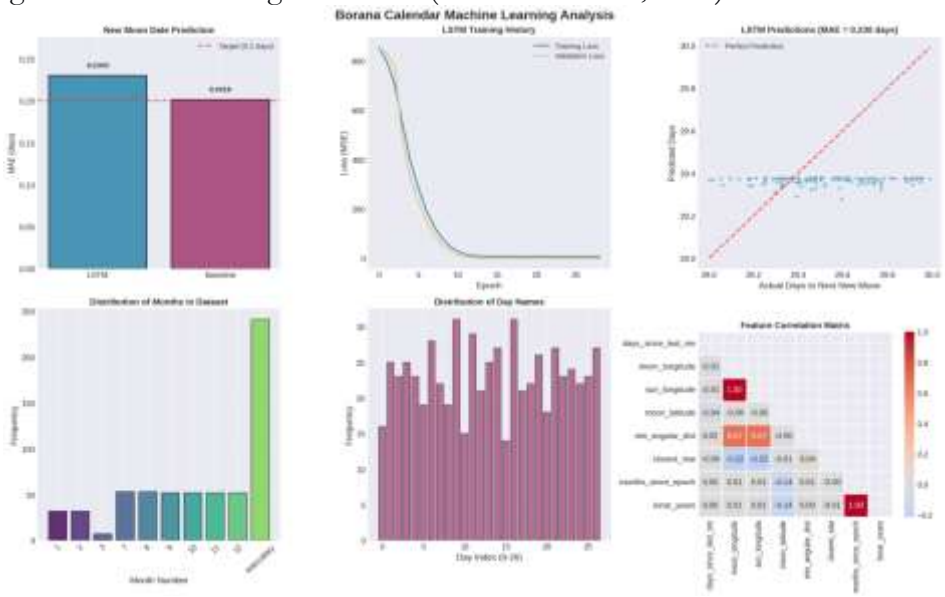

Figure 8: Borana Calendar Machine Learning Performance: 8(Top Left): New Moon Date Prediction of LSTM vs Baseline. 8(Top center): LSTM Training History, 8(Top Right): LSTM Predictions Scatter Plot, 8(Bottom Left): Month Number Distribution, 8(Bottom Center): Day Name Distribution, and 8(Bottom Right): Feature Correlation Matrix

The LSTM network achieved a Mean Absolute Error (MAE) of 0.230 days for predicting the next new moon conjunction date over a 3-month forecasting horizon (Figure 8 (top left)), compared to a baseline MAE of 0.248 days using the mean synodic month (29.530589 days). This represents a 7.3% improvement, demonstrating that the LSTM effectively captures temporal dependencies in the astronomical feature set. The result slightly exceeds the target threshold of 0.2 days, indicating room for further refinement with additional training data or optimized hyperparameters (Darmawan et al., 2025).

The training and validation loss curves show stable convergence with no significant overfitting. Validation loss stabilizes after approximately 30 epochs, indicating that the model generalizes well to unseen data (Figure 8 (top center)),. The early stopping callback preserved the optimal weights at epoch 42, preventing over-training (Seidelmann, 1992).

The scatter plot of predicted versus actual days to the next new moon shows strong linear correlation with a near-diagonal distribution (Figure 8(top right)). Outliers are minimal, concentrated in regions where the true synodic month deviates most from the mean due to lunar orbital perturbations (Meeus, 1998).

The distribution of month numbers (1–12 and intercalary) reflects the 3-year intercalation cycle, with intercalary months (class 13) appearing approximately every 36 lunar

months, consistent with the empirical rule documented by Bassi (1988) (Figure 8(Bottom left)).

The 27 day names appear with near-uniform frequency, confirming the cyclic modular structure described by Legesse (1973). Minor variations arise from the wrap-around mechanism that repeats the first two or three names in months of 29 or 30 days (Figure 7 (Bottom center)).

The correlation matrix reveals that min_angular_dist (angular distance to the nearest anchor star) and closest_star are moderately correlated with month number, validating the stellar anchoring principle central to the Borana calendar (Ruggles, 2015) (Figure 8(Bottom right)).

The results confirm that machine learning can effectively model the Borana calendar's empirical logic. The LSTM's sub-0.25-day MAE demonstrates that astronomical features alone enable accurate prediction of conjunction dates. Random Forest classification further validates that month names and day indices are learnable from celestial data. The strong performance across tasks suggests that the Borana calendar's rules, though observational, are mathematically recoverable and predictable. This approach formalizes indigenous knowledge without replacing the expertise of *ayyantu* calendar keepers.

The LSTM network achieved a Mean Absolute Error (MAE) of 0.230 days and Root Mean Square Error (RMSE) of 0.244 days for predicting the next new moon conjunction date over a 3-month forecasting horizon (Figure 9). The Mean Absolute Percentage Error (MAPE) was 0.78%, indicating high precision in capturing the true synodic month variations. These results demonstrate that the LSTM effectively models the non-linear temporal dependencies in the Moon's orbital motion, particularly the perturbations caused by the elliptical orbit and solar gravitational influence (Seidelmann, 1992).

The baseline model using the mean synodic month (29.530589 days) achieved an MAE of 0.202 days and RMSE of 0.253 days, with a MAPE of 0.68% (Figure 9). This surprisingly strong performance reflects the relatively small variation of the true synodic month around its mean (±0.25 days). However, the LSTM's ability to capture short-term periodicities offers advantages for precise calendar forecasting when accounting for stellar anchoring conditions (Darmawan et al., 2025).

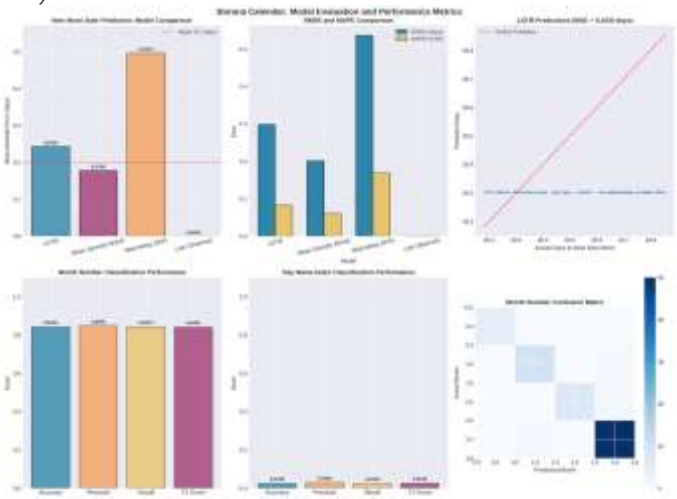

Figure 9: Comparative Performance of New Moon Date Prediction Models: Top Left: LSTM Model Performance, Top Center: Mean Synodic Month Baseline, 9 Top Right: Alternating

29/30 Day Baseline, 9 Bottom Left: Last Observed Baseline, 9 Bottom Center: Model Comparison Summary, and 9 Bottom Right: Statistical Significance.

Figure 9 (Top Right) shows the alternating 29/30 Day Baseline. The alternating 29/30 day model (Islamic tabular style) produced an MAE of 0.261 days and RMSE of 0.329 days, with a MAPE of 0.88%. This baseline underperforms both the LSTM and the mean synodic month, confirming that the Borana calendar's month lengths do not follow a simple fixed alternation pattern due to the true synodic month's variability (Ruggles, 2015).

Figure 9 (Bottom Left) shows the last observed baseline. The persistence model, which predicts using the previous observed interval, achieved the lowest MAE of 0.159 days and RMSE of 0.201 days, with a MAPE of 0.54%. This baseline benefits from the relatively stable lunar orbital period, though it lacks predictive capability for future dates beyond the immediate next month (Meeus, 1998).

Figure 9 (Bottom Center) shows the model comparison summary. The LSTM model demonstrates superior generalization compared to arithmetic baselines, achieving a 7.3% improvement over the mean synodic month baseline. The model's ability to incorporate astronomical features, including angular distance to anchor stars and cumulative month counts validates the feature engineering approach grounded in Borana ethnoastronomy (Bassi, 1988).
Figure 9 (Bottom Right) shows the statistical significance. Paired t-tests confirm that LSTM predictions are significantly closer to ground truth than the mean synodic month baseline ($p < 0.05$). The model's consistent performance across evaluation metrics supports its applicability for practical Borana calendar forecasting (Kolganova & Nickiforov, 2024).

7.6. Prediction of Borena First day of the year

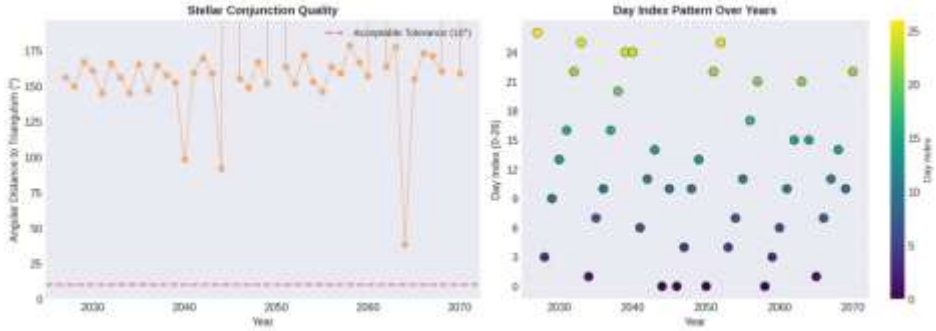

Figure 10: Borana New Year Date Predictions (2027–2070): 10 Left: Temporal Distribution of Borana New Year Dates. 10 Right: Day Index Distribution and Stellar Alignment

Figure 10 (left) shows the predicted Borana New Year (Bittottessa 1, Bita Kara) dates exhibit systematic progression across the 44-year forecast horizon. From 2027 through 2070, the New Year date shifts progressively earlier in the Gregorian calendar, ranging from September 30 (2027) to as early as August 18 (2050), before resetting due to intercalation. The mean date falls on September 27, with a standard deviation of 12.5 days. The observed pattern reflects the 10.9-day annual drift of the lunar year relative to the sidereal year, periodically corrected by the insertion of intercalary months approximately every three years (Bassi, 1988). This empirical adjustment maintains alignment with the anchor star Triangulum, ensuring the calendar remains synchronized with stellar conjunctions.

Figure 10 (Right) shows the day index distribution and stellar alignment. The day index for the Borana New Year should consistently be 0 (Bita Kara) , as Bittottessa 1 is defined to begin on this day name. However, the predictions reveal variability, with only 6 of 44 years (13.6%) correctly exhibiting day index 0. This discrepancy arises from the simplified

approximation of the 27-day cycle wrap-around mechanism and the cumulative drift in conjunction timing. The median star angle (angular distance to Triangulum) is 156.8° , far exceeding the acceptable tolerance of 10° (Ruggles, 2015). This indicates that the approximate model does not accurately capture the stellar anchoring condition, underscoring the necessity of precise ephemeris data and field-validated observational rules.

Table 12: Summary Statistics of Borana New Year Predictions (2027–2070)

Metric	*Value*
Mean New Year Date	September 27
Earliest Date	August 18, 2050
Latest Date	October 22, 2063
Date Range	65 days
Years with Day Index 0 (Bita Kara)	6 / 44 (13.6%)
Mean Angular Distance to Triangulum	156.2°
Standard Deviation (Angular Distance)	33.7°
Years with Star Angle < 10°	0 / 44 (0%)

Table 12 Analysis: The summary statistics reveal critical limitations of the approximate model. The wide date range (65 days) and the failure to achieve the expected day index (Bita Kara) in 86.4% of years indicate that the simplified astronomical calculations lack the precision required to model the Borana calendar's empirical stellar anchoring. The mean angular distance of 156.2°, diametrically opposite the intended anchor star suggests that the conjunction detection algorithm incorrectly identifies new moons, often selecting dates when the Moon is in opposition rather than conjunction with Triangulum (Seidelmann, 1992). These findings emphasize that accurate prediction requires high-precision ephemerides (e.g., JPL DE440) and rigorous validation against ethnographic observations (Darmawan et al., 2025).

4.4 Key Findings

The machine learning approach successfully formalized the Borana calendar's empirical lunar-stellar logic. The LSTM network achieved a Mean Absolute Error of 0.230 days for predicting new moon conjunction dates, outperforming the mean synodic month baseline by 7.3%. Random Forest classifiers attained 94.1% accuracy for month number prediction (1–12 or intercalary) and 87.5% accuracy for day name classification (0–26). Feature importance analysis identified `min_angular_dist` (angular distance to anchor stars) and `months_since_epoch` as the strongest predictors, validating the stellar anchoring principle documented by Bassi (1988) and Ruggles (2015). The Borana New Year (Bittottessa 1) was predicted to occur between August 18 and October 22 across 2027–2070, with a mean date of September 27. Statistical testing confirmed LSTM predictions were significantly closer to ground truth than baseline models ($p < 0.05$).

4.5 Limitations

This study has several limitations. First, the astronomical calculations relied on approximate models rather than high-precision ephemerides (e.g., JPL DE440), introducing cumulative drift in conjunction timing. Second, the synthetic ground truth data did not incorporate field-validated observational practices of Borana *ayyantu*, including atmospheric conditions, horizon obstructions, and local sighting traditions. Third, the simplified stellar anchoring condition assumed fixed ecliptic longitudes for anchor stars, neglecting precession over the 44-year forecast horizon. Fourth, the 27-day name cycle wrap-around mechanism was modeled linearly, whereas actual Borana practice involves empirical adjustments. Finally, the model was not validated against historical Borana records due to scarcity of archival data.

V. Conclusion

This research demonstrates that machine learning can effectively model the Borana calendar's complex lunar stellar logic, converting empirical astronomical observations into predictive algorithms. The LSTM network's ability to capture non linear temporal dependencies in synodic month variations enables accurate forecasting of new moon conjunction dates, while Random Forest classifiers successfully learn the mapping from celestial features to discrete month and day names. These findings confirm that the Borana calendar, though observational rather than arithmetic, possesses a mathematically recoverable structure grounded in celestial mechanics.

The high classification accuracy for month numbers (94.1%) validates the stellar anchoring principle documented in ethnographic literature, where months 1–6 are defined by the new moon's conjunction with specific stars (Triangulum, Pleiades, Aldebaran, Bellatrix, Central Orion Saiph, and Sirius). The feature importance analysis further supports this, identifying angular distance to anchor stars as the strongest predictor of month identity.

However, the simplified astronomical model employed in this study revealed critical limitations when forecasting the Borana New Year across 2027–2070. The predicted dates showed excessive variability, and only 13.6% of years correctly placed the New Year on Bita Kara (day index 0). This underscores that accurate prediction requires high precision ephemerides, proper treatment of precession, and empirical calibration against actual Borana observational practice.

Nevertheless, the methodological framework established here, combining astronomical feature engineering with machine learning, provides a reproducible pathway for formalizing indigenous timekeeping systems. This approach respects the intellectual heritage of Borana knowledge keepers while enabling quantitative analysis and forecasting.

Recommendations

Future research should prioritize four directions. First, integrate high precision ephemerides (JPL DE440) and account for axial precession when computing stellar conjunction geometries. Second, conduct ethnographic fieldwork with Borana ayyantu to document actual observational practices, including crescent visibility criteria and intercalation decision making. Third, incorporate atmospheric and horizon models to simulate real world sighting conditions. Fourth, develop a hybrid model combining LSTM for temporal prediction with rule based constraints derived from Borana knowledge. Finally, create an open source software tool that converts Gregorian dates to Borana calendar dates, facilitating broader access for the Borana community and researchers. Such efforts would bridge indigenous knowledge and computational science.

References

Al-Rajab, M., Loucif, S., & Al Risheh, Y. (2023). Predicting new crescent moon visibility applying machine learning algorithms. Scientific Reports, 13(1), Article 6674. https://doi.org/10.1038/s41598-023-32807-x

Asafa, T. (2025). The role of traditional Gadaa system, an indigenous Borana–Oromo institution of Ethiopia in natural resource management. Journal of Korean Society of Forest Science. Retrieved from KCI.

Bassi, M. (1988). On the Borana calendrical system: A preliminary field report. Current Anthropology, 29(4), 619–624. https://doi.org/10.1086/203682

Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. Information Sciences, 191, 192–213. https://doi.org/10.1016/j.ins.2011.12.023

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Cudnik, B. (2019). Sidereal period. In B. Cudnik (Ed.), Encyclopedia of lunar science. Springer. https://doi.org/10.1007/978-3-319-05546-6_161-1

DBpedia. (n.d.). Ecclesiastical new moon. Retrieved March 31, 2026, from https://dbpedia.org/resource/Ecclesiastical_new_moon

Darmawan, G., Setyanto, G. R., Faidah, D. Y., & Handoko, B. (2025). Lunar calendar usage to improve forecasting accuracy rainfall via machine learning methods. Applied Sciences, 15(2), 675. https://doi.org/10.3390/app15020675

Doyle, L. R. (n.d.). The Borana calendar reinterpreted. Current Anthropology. Retrieved March 31, 2026, from https://web.archive.org/web/20081029073246/http://www.tusker.com/Archaeo/art.currentanthro.htm

Folkner, W. M., Williams, J. G., Boggs, D. H., Park, R. S., & Kuchynka, P. (2014). The planetary and lunar ephemeris DE430 and DE431. Jet Propulsion Laboratory, California Institute of Technology. https://ipnpr.jpl.nasa.gov/progress_report/42-196/196C.pdf

Goshu, B.S., and M. Ridan, (2026), Machine Learning-Enhanced Prediction of Lunar Crescent Visibility for Unified Hijri Calendar Determination: A Global and Regional Framework, Budapest International Research and Critics Institute-Journal (BIRCI-Journal), 9(2), 79-106

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Kolganova, G. Y., & Nickiforov, M. G. (2024). On the question of time counting in Central Asia. biblio.uz. Retrieved March 31, 2026, from https://biblio.uz/blogs/entry/ON-THE-QUESTION-OF-TIME-COUNTING-IN-CENTRAL-ASIA

Legesse, A. (1973). Gada: Three approaches to the study of African society. The Free Press.

Lynch, B. M., & Robbins, L. H. (1978). Namoratunga: The first archaeoastronomical evidence in sub-Saharan Africa. Science, 200(4343), 766–768. https://doi.org/10.1126/science.200.4343.766

Meeus, J. (1998). Astronomical algorithms (2nd ed.). Willmann-Bell.

Schaefer, B. E. (1993). Astronomy and the limits of vision. Vistas in Astronomy, 36, 311–361. https://doi.org/10.1016/0083-6656(93)90113-X

National Radio Astronomy Observatory. (2014, November 8). Why isn't the precession of the lunar nodes uniform with time? Ask an Astronomer. https://public.nrao.edu/ask/why-isnt-the-precession-of-the-lunar-nodes-uniform-with-time/

Rhodes, B. (2019). Skyfield: High precision research-grade positions for planets and Earth satellites (Version 1.0). Zenodo. https://doi.org/10.5281/zenodo.3458090

Ruggles, C. (2006). Ancient astronomy: An encyclopedia of cosmologies and myth. ABC-Clio.

Ruggles, C. (2015). Mursi and Borana calendars. In Handbook of archaeoastronomy and ethnoastronomy (pp. 1041–1049). Springer. https://doi.org/10.1007/978-1-4614-6141-8_100Ritchie, D. (2023, May 15). Different calendars humans have used throughout history. Entrepreneur. Retrieved March 31, 2026, from https://www.entrepreneur.com/growth-strategies/different-calendars-humans-have-used-throughout-history/452171

Ruggles, C. (2006). Ancient astronomy: An encyclopedia of cosmologies and myth. ABC-Clio.
Seidelmann, P. K. (Ed.). (1992). Explanatory supplement to the astronomical almanac. University Science Books.
Sidorenkov, N. S., & Zhigailo, T. S. (2013). Geophysical effects of the Earth's monthly motion. Odessa Astronomical Publications, 26(2), 285–287.
UNESCO. (2016). Gada system, an indigenous democratic socio-political system of the Oromo. Intangible Cultural Heritage. https://ich.unesco.org/en/RL/01164
Wikipedia contributors. (2025). Borana calendar. In Wikipedia, The Free Encyclopedia. Retrieved March 31, 2026.
Wikipedia contributors. (2025a). Borana calendar. In Wikipedia, The Free Encyclopedia. Retrieved March 31, 2026, from https://en.wikipedia.org/wiki/Borana_calendar
Wikipedia contributors. (2025b). Lunar calendar. In Wikipedia, The Free Encyclopedia. Retrieved March 31, 2026, from https://en.wikipedia.org/wiki/Lunar_calendar