Predicting the Visibility of the First Crescent

Tafseer Ahmed[1]

Abstract

This study presents an application of machine learning to predict whether the first crescent of the lunar month will be visible to naked eye on a given date. The study presents a dataset of successful and unsuccessful attempts to find the first crescent at the start of the lunar month. Previously, this problem was solved by analytically deriving the equations for visibility parameter(s) and manually fixing threshold values. However, we applied supervised machine learning on the independent variables of the problem, and the system learnt about the criteria of classification. The system gives precision of 0.88 and recall of 0.87 and hence it treats both false positives and false negatives equally well.

Keyword: crescent visibility, astronomy, supervised learning, feature engineering, ensemble

1 Introduction

Humans are interested in the problem of first sighting of the crescent from ancient times. There were many civilizations that used the lunar or lunisolar calendar, and the first crescent after the sunset marked the beginning of the new lunar month. The following paragraphs explain some fundamental phenomena and terms of astronomy. The details of these terms and phenomena can be found at [1] & [2].

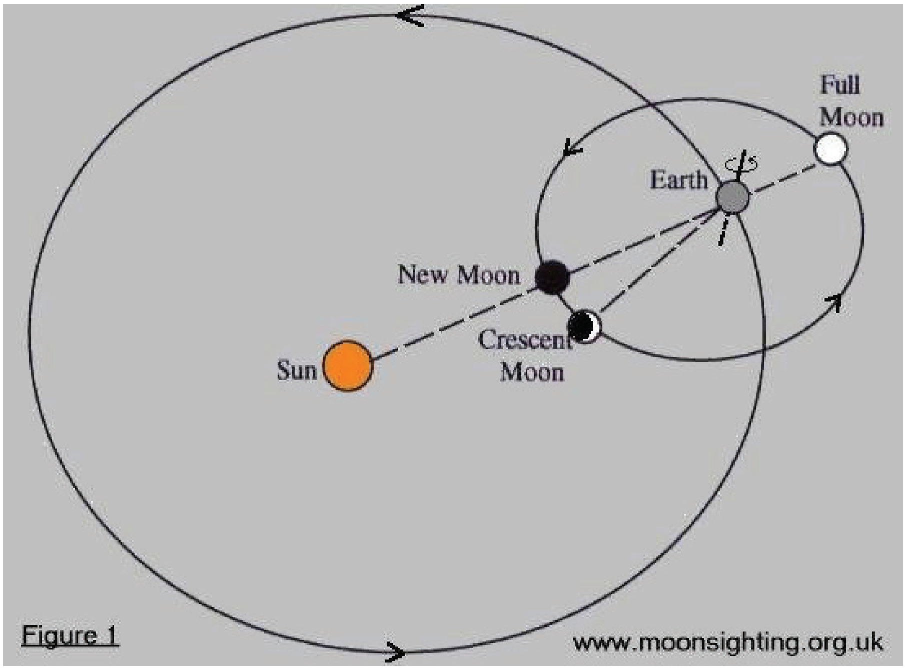

Figure 1: New Moon and Crescent Formation

[1]*Mohammad Ali Jinnah University | tafseer.ahmed@jinnah.edu*

The formation of the crescent is explained in figure 1. The moon rotates around the sun and its relative position with the sun and earth gives the phase of the moon. When the moon is opposite to the sun with respect to earth, we see the full moon. Afterwards, its anticlockwise movement gives waning phases and it becomes invisible after 14 days, as its face towards the earth does not receive light from the sun and it becomes dark.. In the last days of the month, it is visible before the sunrise at the eastern horizon.

When the moon comes between the sun and the earth, or alternatively speaking when sun and moon meet in the sky, we call it sun-moon conjunction or simply conjunction. It is also said that the new moon is born at the conjunction. However, this new moon is not visible to the human eye. The first crescent or the waxing crescent becomes visible above the western horizon after the sunset when it has become sufficiently luminous (by reflecting the light of the sun).

The astronomers know precisely when the (sun-moon) conjunction occurs. The charts are available for the conjunction time of each new moon e.g sun moon conjunction time of each lunar month can be seen at [3]. However, crescent visibility information is not known precisely, because it depends on many factors including human vision, atmospheric conditions, moon's phase and moon's altitude over the horizon. Hence, there exist many models that try to predict the possibility of sighting of the first crescent after the new moon's birth e.g. [3] and [4].

The astronomers focus on this problem for scientific as well social reasons. Many cultures and religions use lunar and lunisolar calendars. They start the new lunar month with the new moon. Many of these use astronomical data to mark the new moon / crescent. The muslim hijri calendar is also a lunar calendar, and the new month start with sight of the crescent [5]. Some countries e.g. Turkey (and its followers) and organizations use the astronomical calculation [4], in place of actual sighting. There are many other calendars on the basis of different visibility or presence criteria of new moon e.g. [6] and [7]. The Ministry of Science and Technology, Pakistan has also issued a criterion and a calendar [8].

However, many communities and countries rely on actual sighting of the crescent (inside or outside the country) to start the new month. Pakistan's authority for moonsighting (Ruet e Hilal Committee) uses crescent sighting models to eliminate improbable claims. In this context, we modeled this problem using machine learning. It is considered as a classification problem having the classes: visible = yes and visible = no. There exists many other works for this prediction, however they are based on astronomical models. According to our knowledge, it is the first attempt to create the crescent visibility model using machine learning.

The remaining paper is organized as follows. Section 2 presents earlier work on the astronomical models of crescent visibility prediction. We mention some limitations of these models, and then present our model in section 3. The section 3 also describes the experiment to evaluate our model. The results of the experiment are presented and discussed in section 4. Section 5 gives description of the further work that can be done to improve the model.

2 Literature Review

The prediction of the first crescent of the month has been an area of interest from the ancient time. Babylonians [9] devised a method to predict visibility of the first moon. Their method involved two features: the elongation and moonset lag time. Table 2 shows that the results of babylonian method are comparable to many modern day methods. However, as the evaluation results show that there were many errors in the prediction. Later astronomers tried to improve the prediction model by adding other factors that affect the prediction of crescent visibility.

In 1911, Maunder [10] proposed that there is a region of brightness around the point of sunset. If the moon is present in that region, then the crescent will not be visible. He proposed the following formula for ARCV (arc of vision). It is calculated by the following equation using the difference in azimuth (DAZ) of the sun and the moon.

$$ARCV = 11 - |DAZ|/20 - (DAZ^2)/100$$

The equation is obtained by learning parameters using observed values of DAZ and ARCV. The crescent is only visible, if its altitude is less than ARCV. Altitude and Azimuth are the coordinates of the objects in the sky [1] [2], and these are explained by the following figure.

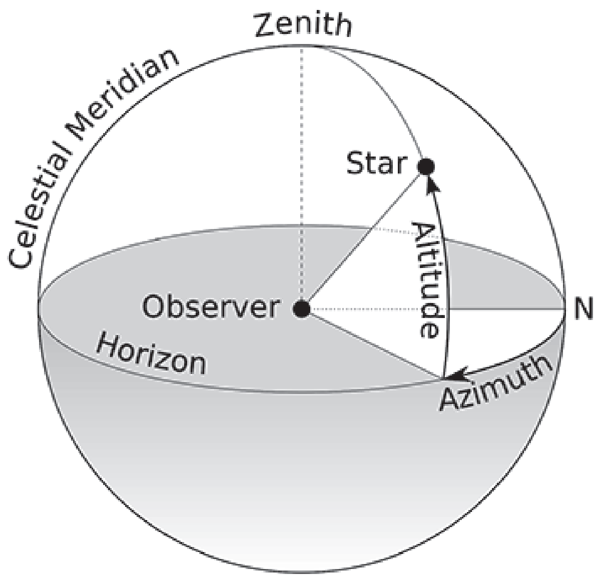

Figure 2: Altitude and Azimuth[1]

In 1930, Schoch gave an alternate equation to calculate ARCV based using the observational data [11]. So, we have alternate methods to calculate ARCV. However, the parameters of both of these equations are derived using a small dataset. In section 3, we collect a larger dataset and propose a method that is able to use the purpose of ARCV without calculating it. The next generation of researchers focused on other factors of crescent visibility. Yallop [12] introduced other derived variables in the calculation. As this paper does not deal with astronomy, so we do not present the equations involved and the reasons for these equations. The calculations can be summarized as:

$$ARCV = f(DAZ), ARCL = f(ARCV, DAZ), W' = f(ARCL), q = f(ARCV, W') \quad (1)$$

Yallop collected 296 observations, and he recorded results of visibility by eye as well as by optical aid (binocular or telescope). By inspecting the observations and corresponding q-values, he presented the following rules.

Table 1: Yallop Visibilaty Rules

Range	Remarks
$q > +0•216$	Easily visible
$+0•216 \geq q > -0•014$	Visible under perfect conditions
$-0•014 \geq q > -0•160$	May need optical aid to find crescent
$-0•160 \geq q > -0•232$	Will need optical aid to find crescent
$-0•232 \geq q > -0•293$	Not visible with a telescope
$-0•293 \geq q$	Not visible

Odeh [13] collected 737 observations to develop a better system of prediction. He derived an equation for the single parameter V, such as V = f(ARCV, W'). On the basis of different values of V, he defined the following four zones for crescent visibility.

- Zone A: visible by naked eyes
- Zone B: visible by optical aid, and it could be seen by naked eyes
- Zone C: visible by optical aid only
- Zone D: not visible even by optical aid

Qureshi [14] worked on the Odeh's dataset, however he removed some observations from this dataset. These removed observations were made in the morning. He derived new equations and then manually defined regions on the basis of threshold values. He compared the results of his method with other methods. The following table is created by using data presented in his evaluation results.

Table 2: Evaluation of different Methods (src: [14])

Criterion	Precision	Recall	F-score	Accuracy
Babylonian	0.63	0.96	0.76	0.75
s-value	0.61	0.94	0.74	0.73
q-value	0.69	0.92	0.79	0.79
Maunder	0.8	0.76	0.78	0.82
Fotheringam	0.87	0.54	0.67	0.79

The above data clearly depicts the precision-recall tradeoff in different methods. The methods with high recall (i.e. low number of false negatives) have low precision (i.e high number of false positives) and vice versa. In this case, a false positive means that the moon is predicted as visible, but it was not actually visible, whereas false negative means that the crescent is predicted as not visible, but it was actually visible. We need a method that minimizes both of these errors. To accomplish this task, we decided to apply machine learning on a bigger dataset. The following section gives the description of our method and its result.

3 Crescent Visibility Prediction

The literature discussed in the above section tells that the first crescent visibility is predicted by deriving different equations on the basis of astronomy based variables, and then manually selecting a cut-off number for the decision of the visibility.

We choose machine learning as an alternate way to model this problem. The advantage of using machine learning is that we do not need to make a theory of crescent visibility and derive different equations. In the presence of training examples, the first crescent visibility can be modeled as a classification task. The classification algorithm can itself learn the importance (e.g. weight) of different astronomy based features. However, for the application of classification algorithms, we need a bigger dataset. Moreover, we need to select (and create) features for the algorithms. The details of these two tasks are presented in section 3A and 3B.

A Dataset Collection

The currently available datasets are not big. Odeh presented 737 observations [13]. Quereshi [14] removed some irrelevant observations from this list and he worked on 463 observations. A recent work has used only 254 observations, of which only 81 are of positive sighting [15]. These examples are not enough to train a machine learning algorithm, hence we searched for more observations.

There are some websites that record the (successful or unsuccessful) attempts to find the first crescent for the islamic month. One of the websites is managed by Islamic Crescent Observation Project (ICOP). The website has records of observations by different professionals and amateurs for 21 lunar years (1419AH to 1441AH) [3]. Every month (of muslim hijri calendar) has a page associated with it. The page has the visibility map and conjunction date and time for the new moon of that month. The details of observations for that month's first crescent are presented on this page.

The earlier years have a free text description of the observations in which the location of the observer, sky conditions and result of the observation are described. However, from 1431AH, we find that the observations are presented in a regular format and each of the following are mentioned for each observation. We term the features of table 3 as raw features. The hijri month and conjunction date are added to the raw features of each observation.

Table 3: ICOP Observation Features

Observation Feature	Possible Values
Time of Observation	After Sunset, ...
Observer Name	
Location	City, State and Country
Sky Condition	Clear, Partially Cloudy, Total Cloudy
Atmosphere	Superb, Clear, Hazy, Very Hazy
Visible By	Eye, Binocular, Telescope, CCD image

The observation results are added as four different features corresponding to Visible by Eye, Visible by Binocular, Visible by Telescope, and Visible by CCD image. The features have binary values, and more than one features can have yes values, as given in the icop observations.

This raw feature dataset has 2592 observations. In this dataset, there were 459 observations in which the sky condition is totally cloudy. As we cannot see the crescent behind the clouds, we removed all the observations having total cloudy sky condition. Similarly, 611 observations were removed that have partly cloudy and not visible (by eye, binocular, telescope, or CCD image) features.

After this filtering, we added astronomy based features to each observation. We termed the resulting feature-set as rich feature-set. The new features added are longitude and latitude of the location, moonset time, sunset time, altitude and azimuth of sun and moon, and moon phase. These features are obtained by using python libraries geopy [16] and pyEphem [17]. The rich feature-set have independent variables related to the first crescent visibility problem. The features used in the machine learning system are described in the next subsection.

B Feature Engineering

An important decision for machine learning is the selection and engineering of the relevant features. We have seen in the literature review that researchers have created equations for ARCV and ARCL etc. on the basis of the variables that are part of our rich feature-set. The weights of some of these equations e.g. (ACRV equation) are learnt using regression. However, this weight learning is performed on a smaller dataset.

We created a rule that only the independent variables and the difference of independent variables are used in our feature-set for machine learning. The contribution of any dependent variables, if relevant, will be devised by the machine learning algorithm. Hence we have the following features in our Machine Learning feature-set.

Table 4: Crescent Features for Machine Learning

Feature	Calculation Method
Age of Moon	conjunction – sunset
Sun Moon Lag	moonset – sunset
Altitude Diff.	moonAltitude - sunAltitude
Azimuth Diff.	| moonAzimuth - sunAzimuth |
MoonPhase	given by pyEphem
Atmosphere	Superb = 1, Clear = 0.85, Hazy=0.6, Very Hazy = 0.35

The value of these features are normalized to a range of 0-1. The atmosphere condition is represented by a number having high value, if the atmosphere is clear. The output variable for this experiment is the binary feature visible by eye. However, in any further work, the other visible by features or their combination can also be used as output variable.

C Experiment

The dataset described in section 3A is transformed into the machine learning features described in table 4. The first 80% of this data (approximately first 8 lunar years) is used as the training data. The remaining data (approximately last 18 months) is used as the test data. We did not split the data randomly into training and test sets, as our current experiment setting is more challenging. The test data is completely unseen. The classifiers have not seen the observation and result of adjacent city or country for the same lunar month, so any overfitted model will not get help in the test.

We used four supervised learning algorithms. Three of these are classification algorithms namely Logistic Regression (LR), Support Vector Machine (SVM), Random Forest Regressor (RFR), and Neural Network (NN). The Neural Network has one hidden layer. All of these algorithms are used by using scikit-learn library [18].

As Random Forest Regressor (RFR) [19] is a regression based technique, it accepts and returns a number as the output. Hence, for RFR, we use visible = yes as 1.0 and visible = no as 0. Then, if the test example gives a result of 0.5 or greater, we consider it as visible = yes. The results of these four supervised learning algorithms are discussed in the next section.

4 Results and Discussion

As described in section 3.3, we applied four supervised learning algorithms on the dataset. The results of applying these models for predicting 304 observations of the test data (last 18 months from 5-1440 AH to 10-1441 AH) are given in table 5.

The fourth row of table 5 gives the evaluation results of the classifier on the basis of Random Forest Regressor (RFR). When the output value is equal or greater than 0.5, then we classify the observation as visible = yes. We used some other values of this threshold (0.5) to improve the recall of the system. We want to reduce the count of false negatives i.e. the cases in which crescent is predicted to be not visible, but it becomes visible to the human eye. Hence, we tried different settings to get a system, given in the fifth row, with better recall.

Table 5: Result

	Precision	Recall	F-score	Accuracy
LR	0.87	0.82	0.84	0.83
SVM	0.88	0.83	0.86	0.84
NN	0.85	0.87	0.86	0.84
RFR (>= 0.5)	0.88	0.88	0.88	0.86
RFR(>= 0.33)	0.83	0.94	0.88	0.86
>=1 positive	0.86	0.9	0.88	0.86
>=2 positive	0.87	0.88	0.88	0.86
>=3 positive	0.88	0.81	0.84	0.83

Similarly, we used ensembles of the four algorithms and created systems that declare visible = yes when at least one, at least two or at least three algorithms give the positive result. The last three rows present results of these ensembles. The >=1 positive gives a good precision, and the highest recall value (among the ensembles). We propose to use it as the best system that balances both the precision and recall.

5 Conclusion and Future Work

We developed a system for predicting first crescent visibility using machine learning algorithms. We created this system by using observations of 11 lunar years. According to our knowledge, it is the first attempt of applying machine learning on first crescent visibility prediction[2]. It shows that the usage of simple features (age of moon, sun_moon_lag, altitude_difference, azimuth_differnce, moon_phase, atmospheric_ondition) can give a good prediction for modeling this astronomy based phenomenon. This method is different from the method used by other researchers who created derived features by using astronomical techniques/models.

We will share the rich feature set of these observations on some dataset sharing platform e.g. Kaggle. It will enable other researchers to apply their models to create an improved system. The system can be improved by adding more observation examples. There is more data available on ICOP [1] and moonsighting [2] website. This data is needed to be manually inserted into the raw dataset. Moreover, a major crowdsource drive of crescent visibility observation is required, as more data (and specially data from different regions of the world will improve the system.) The system has only one feature for the atmospheric conditions, and it is a subjective feature having

[2]*There are some works on searching and detecting the crescent in the sky using computer vision techniques e.g. [20] and [21]. However, our work is different as it predicts the visibility several days or years before the actual time of crescent sighting.*

values superb, clear, hazy and very hazy. In its place, we need empirical features e.g. temperature and humidity etc. as done by [22]. We need to add these into this and future datasets.

References

[1] [1] I. Ridpath, “Oxford Dictionary of Astronomy”, 2nd edition, Oxford University Press, 2012.

[2] W. M. Smart, and R. M. Green. “Textbook on Spherical Astronomy”, revised edition. Cambridge University Press, 1977.

[3] http://www.icoproject.org/res.html?l=en

[4] https://www.moonsighting.com/

[5] M. Ilyas, "Lunar Crescent Visibility Criterion and Islamic Calendar", Quarterly Journal of the Royal Astronomical Society, Vol. 35: 425-461, 1994.

[6] Maskufa, and S. Hidayatulla, “Global Hijriyah Calendar as Challenges Fikih Astronomy”, International Conference on Law and Justice (ICLJ 2017), Indonesia, 2017.

[7] O. Zainon, H. R. Ali and M. F. Abu Hussin, "Comparing the New Moon Visibility Criteria for International Islamic Calendar Concept", 6th International Conference on Space Science and Communication (IconSpace): 144-149, Johor Bahru, Malaysia, 2019.

[8] http://pakmoonsighting.pk/Introduction.aspx

[9] L. J. Fatoohi, F. R. Stephenson, S. S. Al-Dargazelli, The Babylonian First Visibility of the Lunar Crescent: Data and Criterion, Journal for the History of Astronomy 30 (1): 51-72, 1999.

[10] M. Maunder, “On the smallest visible phase of moon”, Journal of the British Astronomical Association, XXI:355-362, 1911.

[11] C. Schoch, “Tafel fur Neulicht”, Ergaenzungsheft zu den Astronomischen Nachrichten, 8(2): B17, 1930.

[12] B. D. Yallop, “A method of predicting the first sighting of new moon”, NAO Technical Note No. 69, HM Nautical Almanac Office, Royal Greenwich Observatory, Cambridge, UK, 1997.

[13] M. S. Odeh, “New criterion for lunar crescent visibility”, Experimental Astronomy 18: 39–64, Springer, 2004.

[14] M. S. Qureshi, “On the comparative study of mathematical models for earliest visibility of the crescent moon and their modification”, Ph.D. Thesis, University of Karachi, 2007.

[15] N. Ahmad et al., “A New Crescent Moon Visibility Criteria using Circular Regression Model: A Case Study of Teluk Kemang, Malaysia”, Sains Malaysiana 49(4)(2020): 859-870, 2020.

[16] https://pypi.org/project/geopy/

[17] B. C. Rhodes, “PyEphem: Astronomical Ephemeris for Python”, Astrophysics Source Code Library, ascl:1112, 2011.

[18] F. Pedregosa, et al., "Scikit-learn: Machine learning in Python", the Journal of machine Learning research 12: 2825-2830, 2011.

[19] T. F. Cootes, et al. ,"Robust and accurate shape model fitting using random forest regression voting", European Conference on Computer Vision. Springer, 2012.

[20] M. Fakhar, et. al., "Lunar Crescent Detection Based on Image Processing Algorithms", Earth, Moon, and Planets, 114.1-2: 17-34, 2014.

[21] K. Alhammadi, et al., "Moon Crescent Tracker", International Conference on Electrical and Computing Technologies and Applications (ICECTA), 2019.

[22] B . E . Schaefer, "Visibility of the lunar crescent", Quarterly Journal of the Royal Astronomical Society 29:511-523, 1988.